THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XX (1957)

EXECUTIVE EDITOR
A. C. STICKLAND

Price to non-Fellows £3 3*s. net* (*Postage* 1*s.* 8*d.*)
Price to Fellows (*one copy*) £1 7*s.* 6*d. net* (*Postage* 1*s.* 8*d.*)

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road
London S.W.7
1957

Printed in Great Britain by John Wright and Sons Ltd.
The Stonebridge Press, Bath Road, Bristol 4

CONTENTS

PAGE

Preface iv

The Theory of Plasma Oscillations in Metals. By S. RAIMES 1

The Theory of Liquid ^{4}He. By J. WILKS 38

High-Energy Protons. By A. E. TAYLOR 86

Theories of Collective Behaviour. By D. TER HAAR 130

K-Mesons and Hyperons : Their Strong and Weak Interactions. By R. H. DALITZ 163

Microwave Properties of Solids. By D. M. S. BAGGULEY and J. OWEN 304

A Review of the Methods of High-Speed Photography. By J. S. COURTNEY-PRATT 379

Photographic Sensitivity. By J. W. MITCHELL 433

Theory and Properties of Solid Argon. By E. R. DOBBS and G. O. JONES . . . 516

Cumulative Author Index, Volumes 16–20 565

Cumulative Subject Index, Volumes 16–20 567

PREFACE

The *Reports on Progress in Physics* is under the general supervision of the Papers Committee of the Physical Society. The Members of the Papers Committee were as follows:

Chairman:—Professor H. Jones.

Honorary Papers Secretary:—Dr. H. H. Hopkins.

Ordinary Members:—The President (Professor N. F. Mott), Dr. M. Blackman, Professor W. E. Burcham, Dr. E. H. S. Burhop, Dr. A. H. Cooke, Professor S. Devons, Professor R. W. Ditchburn, Professor C. Domb, Dr. W. Ehrenberg, Dr. B. H. Flowers, Dr. F. C. Frank, Mr. D. W. Fry, Mr. R. C. Horner, Professor F. Llewellyn Jones, Professor G. O. Jones, Dr. H. G. Kuhn, Professor R. E. Peierls, Mr. E. S. Shire, Dr. R. W. B. Stephens, Dr. C. G. Wynne.

Reports published in the Volume are also being issued as individual articles, at prices depending on their length. Information may be obtained from the offices of the Society.

A. C. STICKLAND,
Secretary–Editor.

THE THEORY OF PLASMA OSCILLATIONS IN METALS

By S. RAIMES
Department of Mathematics, Imperial College, London

CONTENTS

PAGE
§ 1. Introduction 1
§ 2. Previous treatments of the interaction of electrons in metals 3
  2.1 The Sommerfeld and Hartree theories 3
  2.2. The Hartree–Fock theory 6
  2.3. The correlation energy 8
§ 3. Elementary classical theory of plasma oscillations 8
§ 4. The zero-point and excitation energies of a plasma oscillation 10
§ 5. The short-range interaction 11
§ 6. A refinement of the classical theory 13
§ 7. Quantum-mechanical theory 16
§ 8. The correlation energy 22
  8.1. The long-range correlation energy 22
  8.2. The short-range correlation energy 23
§ 9. The effect of plasma oscillations on the energy band theory of metals .. 25
§ 10. Energy losses of fast electrons in penetrating metallic films 27
§ 11. The effects of the periodic lattice field and of the core electrons 30
§ 12. Conclusion 35
Acknowledgments 36
References 36

Abstract. Plasma oscillations in gas discharges have been known and studied for nearly thirty years. The existence of similar oscillations in metals, i.e. longitudinal oscillations of the valence electron gas, analogous to sound waves, was suggested over twenty years ago, but only recently has their full significance been realized. The Coulomb interaction of the electrons can be divided into two parts: a long-range part, whose effect is described by the plasma oscillations; and a residual 'screened' interaction, whose range is about 1 Å. Owing to the large excitation energy of a plasma oscillation, the long-range part can be ignored in many calculations, which explains the success of the independent particle model on which the band theory of metals is based. In the first half of this article, after a critical discussion of previous treatments of the Coulomb interaction, the elements of the plasma theory are made plausible by simple arguments. Then the quantum-mechanical theory of plasma oscillations in a free-electron gas, due to Bohm and Pines, is described, including an account of the calculation of the correlation energy. Experimental evidence of the existence of plasma oscillations, afforded by the energy losses of fast electrons in penetrating thin metallic films, is then discussed. Finally, going beyond the free-electron approximation, a semi-quantitative theory of the effects of the periodic lattice field and of the core electrons is given.

§ 1. INTRODUCTION

A SYSTEM composed of a large number of positive ions and free electrons, with zero total charge, was first called a *plasma* by Langmuir in 1928. The motion of the ions in a plasma, owing to their relatively great mass, can often be neglected in comparison with the motion of the electrons. In fact, it is sufficient for many purposes to regard the plasma as an electron gas moving in a positively charged fluid of uniform density, which is called the *background* of positive charge. The charge density of ions and electrons together is on the average zero, but, owing to thermal fluctuations, the electron density is, in fact, constantly varying. Let us suppose that, as a result of the thermal motion, the electronic charge density in some region is reduced below the average. The positive background is no longer neutralized in that region and the resulting positive charge attracts neighbouring electrons. The tendency is to restore charge neutrality, but the attracted electrons acquire momentum and so overshoot the mark. This produces an excess negative charge which causes the electrons to be repelled outwards again, and so oscillations are set up. These longitudinal oscillations of the electron gas, analogous to sound waves, Langmuir called *plasma oscillations*.

An important example of a plasma, and one which has been the subject of much theoretical and experimental study during the past 30 years, is that occurring in a gas discharge. Tonks and Langmuir (1929) gave a simple theory of such a plasma and derived the frequency of the plasma oscillations; we shall present a similar treatment in § 3. Clearly, a metal provides another example of a plasma, for a metal is a system of positive ions arranged on a crystal lattice, together with valence or conduction electrons which are more or less free to travel throughout the lattice. The sharp distinction between valence electrons and ion-core electrons implied here breaks down somewhat in cases where the cores are large, e.g. in the noble metals and transition metals. We shall return to this question in § 11, but for the present let us bear in mind the simplest model of a metal, as typified by sodium, for which we can certainly say that the core electrons are tightly bound to the nuclei, and the valence electrons, one per atom in the case of sodium, are very nearly free. Unless the contrary is stated, when we refer to the *electrons* of a metal we shall mean the *valence electrons*.

The existence of plasma oscillations in metals was suggested as long ago as 1932, by Steenbeck, who applied the theory of Tonks and Langmuir to the problem of the selective reflection of light from metals. Again, in 1943, Kronig and Korringa treated the energy losses of charged particles passing through metals from the point of view of plasma oscillations, using a hydrodynamical model of a plasma. It is probably true to say that the existence of plasma oscillations, if not recognized as such, has been implicit in much work on the optical properties of metals and on penetration phenomena (as a random sample, see Zener 1933 and Bohr 1948). It is perhaps surprising, therefore, that the full significance of plasma oscillations for the theory of metals was not realized until the appearance of a series of papers on the subject by Bohm and Pines (1951, 1952, 1953, Pines 1953) beginning in 1951. These authors showed that the Coulomb interaction of the electrons in a metal can conveniently be divided into two parts: a long-range part, whose effect is described by the plasma oscillations, and a residual ‘ screened ’ interaction whose range is of the order of 1 Å. We shall see that this not only provides an explanation of the success of the free-electron and more general band theories of metals but also permits the use of perturbation theory in calculations where it was previously impossible, or at least doubtful. It is with this aspect of the plasma theory that we shall principally be concerned in the present article. However, the effects on the optical properties and the energy losses of penetrating particles are of great importance in providing direct experimental evidence of the existence of plasma oscillations, and we shall discuss these topics in §§ 10 and 11.

There is an obvious distinction between a gas discharge plasma and a metallic plasma. In the former the density of electrons is small enough for their motion to be treated by classical mechanics or Maxwell–Boltzmann statistics. The motion of the degenerate electron gas in a metal, however, must properly be treated by quantum mechanics or Fermi–Dirac statistics. None the less, a useful qualitative picture of the plasma oscillations in metals, and even reasonable quantitative results, can be obtained by classical methods, as we shall see. First, however, in § 2 it will be advantageous to review briefly the effects of the electronic interactions, and approximate methods of dealing with them, which were in use before the development of the plasma theory.

The present account is primarily intended for solid state physicists who know something of the electron theory of metals but have not yet had occasion to study plasma oscillations, and perhaps are not well acquainted with methods of calculating the cohesive energy of a metal, where the effects of electronic interactions are of prime importance. The main concern of the article is to make the basis of the theory physically plausible and to indicate the general advantages to be gained from it. The formal quantum mechanics will not be carried out in exhaustive detail, but only the first and most significant steps given. Readers who feel that an even gentler introduction would be desirable may find it in an article by the author (Raimes 1956), which is a very elementary version of part of the present work and involves little mathematics. On the other hand, those who require a more advanced account than is given here, containing details of many applications, will find it in a review article by Pines (1955).

§ 2. PREVIOUS TREATMENTS OF THE INTERACTION OF ELECTRONS IN METALS

The electrons in a metal exert both electrostatic and magnetic forces upon one another. The magnetic forces, however, are generally of negligible magnitude and we shall ignore them; their effect has been considered by Bohm and Pines (1951). The electrostatic or Coulomb interaction of the electrons is of great importance and has always been a stumbling block in the theory of metals. One of the effects of the Coulomb force is to couple, or correlate, the individual electronic motions in such a way as to reduce the probability of any two electrons closely approaching each other. It has been found difficult to incorporate these correlations, which may be called *Coulomb correlations*, in the theory. The main difficulty arises from the fact that the Coulomb interaction is so large and of such long range that satisfactory results cannot be obtained by the standard perturbation methods.

2.1. *The Sommerfeld and Hartree Theories*

In the early Sommerfeld (1928) theory of metals the Coulomb interaction was explicitly ignored, the electrons being assumed to be non-interacting particles contained in a box of constant attractive potential. Furthermore, so long as this model is not applied to the calculation of cohesive energies, and related problems, the depth of the box is immaterial. If periodic boundary conditions are used, the one-electron wave functions are proportional to $\exp(i\mathbf{k}.\mathbf{r})$, where $\mathbf{k}$ is the wave vector, and the density of states in energy $N(\epsilon)$ has the well-known parabolic form shown in figure 1, curve (*a*). At the absolute zero of temperature all one-electron states are occupied up to an energy ζ_0, corresponding to a wave number k_0, above the bottom of the potential box. In calculations which depend mainly upon the density of states, such as those of the electronic specific heats, paramagnetic susceptibilities, and X-ray band widths, the Sommerfeld theory has had considerable success, in spite of its neglect of Coulomb correlations. We shall see that the plasma theory provides an explanation of this fact.

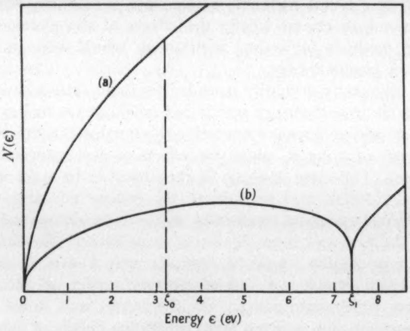


Figure 1. Density of states $N(\epsilon)$ for metallic sodium : (*a*) according to the Sommerfeld and Hartree theories ; (*b*) according to the Hartree–Fock theory at 0°K (all for free electrons). ζ_0 and ζ_1 are the maximum energies at 0°K in the two cases.

In all the following work, with the exception of the semi-quantitative discussion in § 11, we shall assume a free-electron model of a metal, i.e. a model like that of the Sommerfeld theory but which includes the electronic interactions. The electrons are thus *free* only in the sense that the effect of the periodic field of the ionic lattice on the wave functions is neglected, the lattice being replaced by a uniform distribution of positive charge ; they are not free in the sense of *non-interacting*, as in the Sommerfeld theory. The metal is therefore represented by the simple model of a plasma described in the opening paragraph. This is, in fact, the only case for which the theory of plasma oscillations has been worked out in detail at the present time. It is by no means of purely academic interest, however, for we expect several of the alkali metals, sodium in particular, to conform well with such a model, and there is some evidence (see, for instance, Raimes 1950, 1952, 1953) that polyvalent metals with small ion-cores, e.g. magnesium and aluminium, may sometimes be approximately represented in this way.

This model is easily applied to the calculation of the cohesive energies of actual metals, at least ‘ free-electron ’ metals, by suitably choosing the depth of the potential box. In their initial work on sodium, for instance, Wigner and Seitz (1933) effectively took this to be the difference between the energy of the valence electron in a free atom and the ‘ ground state ’ energy ϵ_0 of an electron in the metal. The latter was obtained approximately by solving the Schrödinger equation

$$\frac{d^2\psi}{dr^2}+\frac{2}{r}\frac{d\psi}{dr}+\frac{2m}{\hbar^2}[\epsilon_0-V(r)]\psi=0, \qquad \ldots\ldots(2.1)$$

subject to the boundary condition

$$\left(\frac{d\psi}{dr}\right)_{r=r_s} = 0. \qquad \ldots\ldots(2.2)$$

Here r_s is the radius of an 'atomic sphere', i.e. a sphere whose volume is equal to the mean volume per atom of the metal, and $V(r)$ is the ion-core potential, assumed to be the same as in a free atom. The effect of the periodic lattice field on the kinetic energy of the electrons can often be included by replacing the electronic mass m in the kinetic energy formula by some 'effective mass' m^*, but this is not so easy in energy terms arising from the electronic interactions. Except in § 11 we shall neglect the difference between m and m^*, a procedure well justified in the case of sodium, for which $m^* = 0{\cdot}98m$ (Brooks 1953).

For convenience we may take our zero of energy to be the bottom of the potential box, so that we shall only be concerned with kinetic energy and electronic interaction terms. In order that our expressions may apply to suitable metals of any valency we shall introduce another radius r_e, that of a sphere whose volume is the mean volume *per electron*. Thus, if n is the number of electrons per unit volume, we have

$$\frac{1}{n} = \frac{4\pi}{3} r_e^3. \qquad \ldots\ldots(2.3)$$

If Z is the valency, the relation between r_e and the 'atomic radius' r_s, defined above, is

$$r_e = Z^{-1/3} r_s. \qquad \ldots\ldots(2.4)$$

For monovalent metals, of course, $r_e = r_s$.

The earliest theory which explicitly included to some extent the Coulomb interaction of the electrons was that of Hartree (1928). Coulomb correlations are again ignored, however, and only the average effect of the electrons upon one another considered, i.e. the electrostatic field in which any electron moves is taken to be the ionic field plus the field due to the average charge distribution of the other electrons. In a metal the electron density is so great that it makes no difference, in fact, if the charge distribution of *all* the electrons is used instead of just that of the *other* electrons. In our free-electron model the average charge density is uniform so that the net effect is just to increase the energy of each electron by the same amount. In order to conform with the energy calculation in § 7 this average potential energy must also be included in the depth of the potential box—in fact, the effect of the average electronic charge distribution is exactly cancelled by the assumed uniform distribution of positive charge. The energy of an electronic state with wave vector $\mathbf{k}$ is then just the kinetic energy

$$\epsilon_{\mathrm{H}}(\mathbf{k}) = \hbar^2 k^2/2m, \qquad \ldots\ldots(2.5)$$

and the density of states in the Hartree theory is, for our model, exactly the same as in the Sommerfeld theory. The average energy per electron is easily found to be

$$E_{\mathrm{H}} = \tfrac{3}{5}\zeta_0 = \frac{3\hbar^2}{10m}\left(\frac{9\pi}{4}\right)^{2/3}\frac{1}{r_e^2}. \qquad \ldots\ldots(2.6)$$

It will prove convenient to use the *rydberg* (ryd) as unit of energy and the *Bohr unit* (B.u.) as unit of length, defined by

$$1 \text{ rydberg} \quad = me^4/2\hbar^2 = 13{\cdot}60 \text{ ev},$$

$$1 \text{ Bohr unit} = \hbar^2/me^2 = 0{\cdot}5292 \text{ Å}.$$

For brevity we shall refer to these as 'atomic units', although this is not customary. In atomic units, then, we have

$$E_{\mathrm{H}} = \frac{2{\cdot}21}{r_e^2} \text{ ryd.} \qquad \ldots\ldots(2.7)$$

The Hartree theory thus has the same success as the Sommerfeld theory in calculations depending upon the density of states. When due allowance is made for the depth of the potential box, however, the Hartree theory is found to fail completely in cohesive energy calculations (Wigner and Seitz 1934); in fact, it gives no cohesion. The reason for this is simply that Coulomb correlations have been ignored.

2.2. *The Hartree–Fock Theory*

It is well known that the Hartree method is equivalent to the use of a total wave function which is a single product of one-electron functions. An improvement on this is provided by the Fock (1930), or Hartree–Fock, method which uses a total wave function consisting of a determinant of one-electron functions, the wave function thus being antisymmetric in the electronic coordinates, as required by the Pauli principle. For future reference we shall give this determinant, which we call D_0, explicitly; for unit volume of metal, containing n electrons, it is

$$D_0 = \frac{1}{\surd(n!)} \begin{vmatrix} \psi_1(\mathbf{r}_1') & \psi_1(\mathbf{r}_2') & \cdots & \psi_1(\mathbf{r}_n') \\ \psi_2(\mathbf{r}_1') & \psi_2(\mathbf{r}_2') & \cdots & \psi_2(\mathbf{r}_n') \\ \vdots & \vdots & & \vdots \\ \psi_n(\mathbf{r}_1') & \psi_n(\mathbf{r}_2') & \cdots & \psi_n(\mathbf{r}_n') \end{vmatrix}, \qquad \ldots\ldots(2.8)$$

the factor $1/\surd(n!)$ normalizing the function in unit volume. The primes indicate that spin coordinates are included; the one-electron functions, again normalized in unit volume, have the form

$$\psi(\mathbf{r}') = \exp(i\mathbf{k}\,.\,\mathbf{r})\,\alpha(\eta), \qquad \ldots\ldots(2.9)$$

where $\alpha(\eta)$ is a spin function. In the ground state of the system, which is the state whose energy we require, and assuming that the metal is not ferromagnetic, each of the orbital states $\exp(i\mathbf{k}\,.\,\mathbf{r})$ appears twice in D_0, associated with opposite spin functions.

The use of a wave function of this type has the effect of lowering the probability of two electrons *with parallel spins* closely approaching one another (see, for example, Seitz 1940, Chap. 6). One may describe the situation roughly but graphically by saying that during its motion any electron is surrounded by a spherical *hole* in the distribution of electrons with parallel spins. This is sometimes called the *Fermi hole* or *exchange hole* and its radius is approximately r_e. We thus see that in the Hartree–Fock method the motions of two electrons with parallel spins are correlated to some extent, although this correlation is an ‘accidental’ one due to the Pauli principle rather than to the Coulomb repulsion. The motions of two electrons with antiparallel spins are still completely uncorrelated.

In the Hartree–Fock theory (see, for example, Wilson 1953, p. 79) the energy of a state with wave vector **k**, referred to the same zero as in the Hartree theory, is

$$\epsilon_{\mathrm{HF}}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m} - \frac{e^2}{\pi}\left[2k_0 + \frac{k_0{}^2 - k^2}{k}\ln\left|\frac{k_0+k}{k_0-k}\right|\right], \qquad \ldots\ldots(2.10)$$

where k_0 is the radius of the Fermi sphere. The second term is the *exchange energy* associated with wave vector **k**. The presence of this term has an unfortunate effect upon the density of states (see Raimes 1954 b, c). If the latter is calculated in the usual way, by means of the formula

$$N(\epsilon) = k^2(2\pi^2\, d\epsilon/dk), \qquad \ldots\ldots(2.11)$$

a distribution of the type shown in curve (*b*) of figure 1 results. It is immediately apparent that agreement with experiment must be completely destroyed in cases, depending upon the density of states, where the Hartree theory was successful. In particular, the band width is now much too large (Jones 1954), being about twice the experimental x-ray band width. Also the calculated electronic specific heat is very much affected since it depends upon the density of states near the Fermi surface, i.e. $k = k_0$, and in the Hartree–Fock theory the density falls rapidly to zero at the Fermi surface and then rises again. In fact, the electronic specific heat at low temperatures T is found to be proportional to $T/\ln T$ (Bardeen 1936, cf. Wohlfarth 1950) instead of proportional to T as observed and as predicted by the Hartree theory. Perhaps it should be made clear that the Hartree–Fock distribution shown in figure 1 is that which applies at the absolute zero of temperature; as the temperature is raised the shape of the curve will change, unlike that of the Hartree theory, owing to the dependence of the exchange energy upon the *occupation* of states.

The reason for these discrepancies is simply that ‘exchange’ effects are exaggerated by the neglect of proper Coulomb correlations. In the plasma theory the long-range Coulomb correlations are taken into account before the Pauli principle comes into play and, as will be seen in § 9, the resulting exchange energy and related effects are much smaller. None the less, even the ‘accidental’ inclusion of parallel spin correlations has a salutary effect upon the total energy, which in the Hartree–Fock theory, using atomic units, is *per electron*

$$E_{\mathrm{HF}} = \frac{2{\cdot}21}{r_e{}^2} - \frac{0{\cdot}916}{r_e}\ \text{ryd.} \qquad \ldots\ldots(2.12)$$

The lowering of the energy brought about by the second term, the average exchange energy, is a great improvement on the simple theory, but unfortunately it is still not enough to give anything like the right cohesive energy. What is still lacking is an energy term, comparable in size with the cohesive energy, which is due principally to antiparallel spin correlations; it is called the *correlation energy*.

2.3. *The Correlation Energy*

Owing to the unfortunate dichotomy of correlations according to spin the terminology in this part of the subject is apt to be confusing. We therefore define the correlation energy formally as follows :

The correlation energy is the difference between the Hartree–Fock energy and the energy obtained when Coulomb correlations are correctly taken into account.

Wigner (1934, 1938) obtained the following approximate formula for the correlation energy per electron, which we denote by W, of a free-electron gas :

$$W = -\frac{0{\cdot}58}{r_e + 5{\cdot}1}\,\text{ryd}. \qquad \ldots\ldots(2.13)$$

This formula, whose accuracy Wigner estimated to be within 20%, has been successfully used in many cohesive energy calculations, particularly for the alkali metals (see, for example, Brooks 1953). The general outline of Wigner's method has often been described (see Seitz 1940, p. 343 ; Pines 1955, p. 374) and no useful purpose would be served by reproducing it here. The detailed working is sufficiently obscure, however, for opinion to have sometimes been divided upon whether the formula gives even the right order of magnitude (see Slater 1953, cf. Raimes 1954 a). It will be shown in § 8 that the plasma theory provides a straightforward verification of Wigner's result.

It should be mentioned here that Macke (1950) has also calculated the correlation energy of a free-electron gas and obtained a result similar to Wigner's. Macke treated the whole Coulomb interaction of the electrons as a perturbation. It is well known that the second-order energy correction diverges in this case, but by proceeding to higher orders Macke was able to obtain a convergent result.

Neither of these methods of treating correlations provides a straightforward explanation of the anomalous density of states of the Hartree–Fock theory. In the plasma theory, however, the Coulomb interaction is divided naturally into long-range and short-range parts, and the description of the effect of the former in terms of plasma oscillations provides the necessary explanation. Furthermore it is found that, owing to this division, second-order perturbation theory can be used in calculating the correlation energy.

§ 3. ELEMENTARY CLASSICAL THEORY OF PLASMA OSCILLATIONS

For the present let us ignore possible effects on the plasma oscillations of the degeneracy of the electron gas in a metal, returning to this in § 7. We shall consider the metallic system to be represented by a simple plasma and treat the motion of this by classical electrodynamics.

Let us consistently denote by n the average density of electrons, so that, if $-e$ is the charge of an electron, the average electronic charge density is $-en$. The uniform charge density of the positive background is thus en. Also let us denote by n' the *actual* density of electrons, which is a function of position and of time ; it is, in fact, at any point an instantaneous average over a region small compared with the whole volume of the metal but still containing a large number of electrons.

The excess positive charge density is then $e(n-n')$ and Poisson's equation gives

$$\operatorname{div}\mathcal{E} = 4\pi e(n-n'), \qquad \ldots\ldots(3.1)$$

where $\mathcal{E}$ is the electric field.

Now suppose that the varying density n' is due to a displacement $\boldsymbol{\xi}$ of the electron gas, this being a 'drift' displacement, again a function of position and of time, superimposed upon the random thermal motion of the electrons. The particle current density is therefore $n'\dot{\boldsymbol{\xi}}$, and satisfies the *equation of continuity*

$$-\frac{\partial n'}{\partial t} = \operatorname{div}(n'\dot{\boldsymbol{\xi}}). \qquad \ldots\ldots(3.2)$$

We shall assume that $\boldsymbol{\xi}$ is always small, so that $n-n'$ is small; this being equivalent to the assumption that the plasma oscillations are of small amplitude. In this case div $(n'\dot{\boldsymbol{\xi}})$ can be replaced to a sufficiently good approximation by $n \operatorname{div} \dot{\boldsymbol{\xi}}$ in (3.2) and the resulting equation integrated with respect to time, giving

$$n-n' = n \operatorname{div} \boldsymbol{\xi}. \qquad \ldots\ldots(3.3)$$

We therefore have, from equations (3.1) and (3.3),

$$\operatorname{div} \boldsymbol{\mathscr{E}} = 4\pi en \operatorname{div} \boldsymbol{\xi}, \qquad \ldots\ldots(3.4)$$

or

$$\boldsymbol{\mathscr{E}} = 4\pi en\boldsymbol{\xi}. \qquad \ldots\ldots(3.5)$$

No additive function of position is required here, since $\boldsymbol{\mathscr{E}}$ must be zero everywhere when $\boldsymbol{\xi}$ is zero everywhere.

Finally, the equation of motion of an electron in the field $\boldsymbol{\mathscr{E}}$ is

$$m\ddot{\boldsymbol{\xi}} = -e\boldsymbol{\mathscr{E}}. \qquad \ldots\ldots(3.6)$$

Hence, from (3.5) and (3.6), we have

$$m\ddot{\boldsymbol{\xi}} + 4\pi e^2 n\boldsymbol{\xi} = 0. \qquad \ldots\ldots(3.7)$$

This is the equation of simple harmonic motion and tells us that the electron gas oscillates with angular frequency

$$\omega_{\mathrm{p}} = (4\pi e^2 n/m)^{1/2}, \qquad \ldots\ldots(3.8)$$

which is called the *plasma frequency*. For a typical gas discharge ω_{p} is of the order of 10^{10} sec^{-1}, whereas for a metal it is of the order of 10^{16} sec^{-1}.

Of course, in the foregoing discussion the effects of the thermal motion of the electrons have been neglected, except in so far as thermal fluctuations will cause a departure from uniform electronic density and so initiate the plasma oscillation. A more complete classical treatment (Bohm and Pines 1952), only the first steps of which will be given in § 6, shows that the angular frequency of oscillation ω is ω_{p} for very long waves, but increases as the wavelength λ decreases. The dispersion relation is found to be approximately

$$\omega^2 = \omega_{\mathrm{p}}{}^2 + k^2\langle v_i{}^2\rangle_{\mathrm{av}}, \qquad \ldots\ldots(3.9)$$

where $k(=2\pi/\lambda)$ is the wave number of the plasma wave, $\mathbf{v}_i$ is the velocity of the ith electron, and the average is over all the electrons. The dispersion relation resulting from the quantum-mechanical treatment, equation (7.22), differs from this only by terms of higher order in k. It is of interest to note here that Wolff (1953) has obtained this same dispersion relation using a semi-classical theory of plasma oscillations together with the Hartree theory of a free-electron gas.

In fact, the dispersion is not very great, ω never differing by more than a few per cent from ω_p, because there is a natural upper limit to k, i.e. there is a lower limit to λ, for wavelengths below which plasma oscillations cannot occur. That this must be so follows at once from the corpuscular nature of the electron gas. It is meaningless to speak of organized oscillations of the latter with wavelengths smaller than the average interelectronic distance. According to the elementary Debye theory of specific heats a solid may be assumed to vibrate thermally like a continuous elastic medium so long as the wavelength of vibration is greater than about *twice* the interatomic spacing (see, for example, Slater 1939, Chap. 14). By analogy we might expect the smallest wavelength of plasma oscillation, λ_c say, to be about twice the average interelectronic distance. A rough estimate of the latter is $2r_e$, where r_e is defined by equation (2.3). Thus λ_c is approximately $4r_e$, and the largest value of k, which we denote by k_c, is roughly given by

$$k_c = \frac{2\pi}{\lambda_c} \simeq \frac{\pi}{2r_e}. \qquad \ldots\ldots(3.10)$$

For sodium, with $r_e \simeq 2$ Å, we find $\lambda_c \simeq 8$ Å and $k_c^{-1} \simeq 1{\cdot}3$ Å, whereas the quantum-mechanical calculation of Bohm and Pines, which we shall describe in § 7, gives $k_c^{-1} = 1{\cdot}56$ Å. Of course the above method of estimation is very crude; it gives a reasonable lower bound to λ, but we have no means of knowing whether plasma oscillations persist down to this wavelength, for the method fails to take into account the tendency of the thermal motion of the electrons to disrupt the organized oscillation of the electron gas. However, for our present purposes it is sufficient to recognize that a maximum value of k, namely k_c, exists, and that k_c^{-1} is of the order of 1 Å.

§ 4. THE ZERO-POINT AND EXCITATION ENERGIES OF A PLASMA OSCILLATION

We shall now make an elementary transition to quantum theory. Let us disregard dispersion for the moment and assume that the plasma oscillations in a metal can be represented by a finite set of harmonic oscillators with angular frequency ω_p. The number of oscillators will depend upon the value of k_c. According to elementary quantum theory the energy of such an oscillator must have one of the values given by the expression $(\mathcal{N}+\frac{1}{2})\hbar\omega_p$, where $\mathcal{N}$ is a positive integer or zero. The state with $\mathcal{N}=0$, or ground state, has energy $\frac{1}{2}\hbar\omega_p$, which is the ***zero-point energy*** of the oscillator. In the classical theory a plasma oscillation would be damped, owing to collisions, etc., and would eventually die out, but according to the quantum theory it will never be completely dissipated but will always be present with at least the energy $\frac{1}{2}\hbar\omega_p$. The energy required to raise the oscillator from one state to that of next higher energy, the ***excitation energy***, is $\hbar\omega_p$. The success of the plasma theory as a computational tool depends crucially upon the magnitude of $\hbar\omega_p$, which we shall now calculate.

Equations (2.3) and (3.8) give

$$\omega_p^2 = \frac{4\pi e^2 n}{m} = \frac{3e^2}{m r_e^3}. \qquad \ldots\ldots(4.1)$$

Changing to the atomic units defined in § 2, we obtain

$$\hbar\omega_p = \frac{2\sqrt{3}}{r_e^{3/2}}\text{ ryd.} \qquad \ldots\ldots(4.2)$$

For sodium, with $r_e = 4$ B.u., this gives $\hbar\omega_p = 0{\cdot}433$ ryd $= 5{\cdot}9$ ev. In general, for all metals to which the free-electron approximation might conceivably apply, $\hbar\omega_p$ lies between 3 and 25 ev (Pines 1955, Table X).

It is of some interest to compare $\hbar\omega_p$ with ζ_0, the maximum kinetic energy of any electron at 0°K, according to the Sommerfeld–Hartree theory. We have

$$\zeta_0 = \frac{\hbar^2}{2m}\left(\frac{9\pi}{4}\right)^{2/3}\frac{1}{r_e^2}, \qquad \ldots\ldots(4.3)$$

or, in atomic units,

$$\zeta_0 = 3{\cdot}684/r_e^2\text{ ryd.} \qquad \ldots\ldots(4.4)$$

Thus

$$\hbar\omega_p = 0{\cdot}941\, r_e^{1/2}\zeta_0 \qquad \ldots\ldots(4.5)$$

when r_e is measured in B.u.

Now $0{\cdot}941 r_e^{1/2} > 1$ for all metals, so that $\hbar\omega_p > \zeta_0$. In fact $\hbar\omega_p$, the excitation energy of a plasma oscillation, is greater than the kinetic energy of any electron at normal temperatures. It is therefore very much greater than the thermal energy which any electron is able to give away. This means that temperature excitation can usually be neglected and the oscillations will remain in their ground states unless excited by some other method, such as that described in § 10.

The foregoing remark is a very important one. The utility of the plasma theory depends upon it, for it means that the plasma oscillations take no active part in many electronic processes and may often be ignored. A plasma oscillation is an organized motion of a large number of electrons and does not occur with wavelength smaller than a few ångström units. It may be inferred from this that the *long-range* part of the Coulomb interaction is responsible for the plasma oscillations, so that in calculations where the latter may be ignored the former may also be ignored. We shall see that the remaining interaction has an effective range of the order of 1 Å, so that reasonable results might often be expected by treating the electrons as non-interacting. This explains the success of the independent particle approximation, upon which the band theory of metals is based.

The method of separating the Coulomb interaction into long-range and short-range parts, and the effects of each, will be explained in the following sections.

§ 5. THE SHORT-RANGE INTERACTION

The Coulomb interaction of the electrons is represented in the Hamiltonian, equation (7.1), by

$$\tfrac{1}{2}\sum_{i,j\neq i}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|}, \qquad \ldots\ldots(5.1)$$

where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the position vectors of the ith and jth electrons respectively. It will be seen in § 7 that the quantum-mechanical treatment of the plasma oscillations has the effect of separating this expression into two parts, essentially with small values of $|\mathbf{r}_i-\mathbf{r}_j|$ and with large values of $|\mathbf{r}_i-\mathbf{r}_j|$ respectively, i.e. the short-range and long-range parts discussed above. The method of effecting this separation is to express (5.1) as a Fourier series, valid within a box of finite volume, and this series is split into two parts consisting of terms with wave number k greater than k_c and of terms with k less than k_c respectively, k_c being the maximum wave number of plasma oscillation introduced in § 3. The terms with k less than k_c are eventually incorporated, after a canonical transformation, in a set of harmonic oscillator Hamiltonians which represent the plasma oscillations. The terms with k greater than k_c remain unchanged after the transformation and represent a residual short-range *screened* interaction among the electrons, as we shall now see.

The potential energy of two electrons, a distance r apart, i.e. a single term of (5.1), can be expressed as follows:

$$\frac{e^2}{r}=\frac{e^2}{2\pi^2}\int\frac{1}{k^2}\exp(i\mathbf{k}.\mathbf{r})\,d\mathbf{k}=\frac{e^2}{\pi}\int_0^\infty dk\int_0^\pi \exp(ikr\cos\theta)\sin\theta\,d\theta. \qquad \ldots\ldots(5.2)$$

This is a straightforward mathematical identity whose correctness may be established by carrying out the integrations on the right-hand side. We are here using a three-dimensional Fourier integral representation, for mathematical convenience, rather than a Fourier series in a finite volume; this does not affect the result because the series is always ultimately approximated by an integral of the above type. In accordance with the method outlined, we divide the integral with respect to k into two ranges, 0 to k_c and k_c to ∞. The latter gives the potential energy of the short-range interaction between the two electrons, i.e.

$$\frac{e^2}{\pi}\int_{k_c}^\infty dk\int_0^\pi \exp(ikr\cos\theta)\sin\theta\,d\theta$$
$$=\frac{2e^2}{\pi}\int_{k_c}^\infty\frac{\sin kr}{kr}\,dk=\frac{e^2}{r}F(k_c r), \qquad \ldots\ldots(5.3)$$

where

$$F(k_c r)=1-\frac{2}{\pi}\int_0^{k_c r}\frac{\sin x}{x}\,dx. \qquad \ldots\ldots(5.4)$$

The total potential energy due to the short-range interaction, which we denote by $H_{\text{s.r.}}$, is thus

$$H_{\text{s.r.}}=\tfrac{1}{2}\sum_{i,j\neq i}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|}F(k_c|\mathbf{r}_i-\mathbf{r}_j|). \qquad \ldots\ldots(5.5)$$

The graph of the function $F(k_c r)$ is shown in figure 2, and the function $\exp(-k_c r)$ is also shown for comparison. For values of $k_c r$ less than 2 the functions are very similar, falling rapidly from unity at the origin to nearly zero. The damped oscillations which occur in $F(k_c r)$ further on are of small amplitude, and are negligible when multiplied by e^2/r—they have no profound physical significance, but are due to the assumption that the plasma oscillations cease abruptly at $k=k_c$, which is rather an oversimplification of the problem. We see, then, that the short-range interaction, that which remains when the long-range interaction has been accounted for by the plasma oscillations, is a screened Coulomb interaction and the effective range, or *screening distance*, is roughly k_c^{-1}, i.e. one or two ångström units.

A screening of the charge of an electron might be expected simply as a result of the polarization of the surrounding plasma, i.e. neighbouring electrons are repelled leaving an excess positive charge around the given electron. A simple application of the Thomas–Fermi method to this case (Mott and Jones 1936, p. 87, Pines 1955, p. 381) gives for the screened potential

$$\frac{e}{r}\exp(-r/b), \qquad \ldots\ldots(5.6)$$

of similar form to that found above, but with screening distance b only about one half of k_c^{-1}. It is, in fact, somewhat misleading to think of the short-range interaction in this way. It must be remembered that in our problem the screening is not due to a static polarization of the plasma, but is a dynamic effect due to a re-description of the motion of the electrons in terms of plasma oscillations plus short-range collisions; the whole Coulomb interaction of the electrons is still present but it has conveniently been divided into two parts.

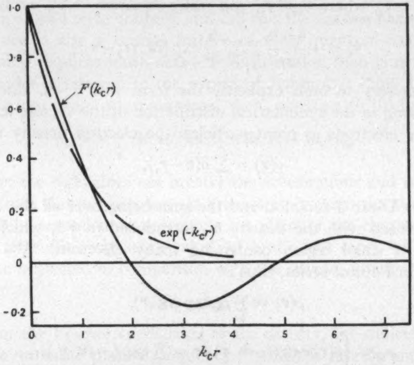


Figure 2. The full curve represents the function $F(k_c r)$, equation (5.4), and the broken curve the function $\exp(-k_c r)$.

The development so far has, of course, been purely tentative. It would indeed be difficult to accept without the help of evidence from experimental results and more detailed calculation, which will concern us in the remaining sections.

§ 6. A REFINEMENT OF THE CLASSICAL THEORY

Before going on to discuss the quantum-mechanical theory of plasma oscillations it will prove helpful to present a slightly more sophisticated form of the classical theory (Bohm and Pines 1952), which introduces several conceptions useful in the quantum theory and shows more clearly the competition between the organized oscillations and thermal motion of the electrons.

For computational convenience let us assume that the metal is in the form of a cube of unit volume. In our free-electron model we have, then, n electrons in a cubic box of unit volume containing also a uniform distribution of positive charge just sufficient to neutralize the electronic charge. The potential energy due to the Coulomb repulsion between the ith and jth electrons is $e^2/|\mathbf{r}_i-\mathbf{r}_j|$, and this may be expanded in Fourier series within the box, as follows:

$$\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} = 4\pi e^2 \sum_{\mathbf{k}}{}' \frac{1}{k^2} \exp[i\mathbf{k}.(\mathbf{r}_i-\mathbf{r}_j)], \qquad \ldots\ldots(6.1)$$

where $\mathbf{k} = 2\pi(n_1\mathbf{a}_1+n_2\mathbf{a}_2+n_3\mathbf{a}_3)$, $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ being unit vectors along three mutually orthogonal edges of the cube, and n_1, n_2, n_3 being integers. The prime on the summation sign indicates that $\mathbf{k} = \mathbf{0}$ is excluded; we shall see below that this is justified by the presence of the uniform background of positive charge.

The total Coulomb force exerted on the ith electron by all the other electrons is, then,

$$-\sum_{j\neq i} \operatorname{grad}_i \left(\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|}\right). \qquad \ldots\ldots(6.2)$$

Equating this to $m\dot{\mathbf{v}}_i$, where $\mathbf{v}_i = \dot{\mathbf{r}}_i$, and using equation (6.1), we find

$$\dot{\mathbf{v}}_i = -\frac{4\pi e^2 i}{m} \sum_j \sum_k{}' \frac{\mathbf{k}}{k^2} \exp[i\mathbf{k}.(\mathbf{r}_i-\mathbf{r}_j)]. \qquad \ldots\ldots(6.3)$$

Here it is unnecessary to omit explicitly the term with $j = i$, since this is automatically zero owing to the symmetrical distribution of the vectors $\mathbf{k}$.

Regarding the electrons as point particles, the electron density may be written

$$\rho(\mathbf{r}) = \sum_i \delta(\mathbf{r}-\mathbf{r}_i), \qquad \ldots\ldots(6.4)$$

$\delta(\mathbf{r}-\mathbf{r}_i)$ being the Dirac δ-function and the sum being over all the electrons; this must not be confused with the density n', introduced in § 3, which is, of course, an average over a small region containing many electrons. We may formally expand $\rho(\mathbf{r})$ also in Fourier series, thus:

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}} \rho_{\mathbf{k}} \exp(i\mathbf{k}.\mathbf{r}), \qquad \ldots\ldots(6.5)$$

where
$$\rho_{\mathbf{k}} = \int \rho(\mathbf{r}) \exp(-i\mathbf{k}.\mathbf{r})\, d\mathbf{r} = \sum_i \exp(-i\mathbf{k}.\mathbf{r}_i). \qquad \ldots\ldots(6.6)$$

Now $\rho_{\mathbf{0}}$, the Fourier coefficient with $\mathbf{k} = \mathbf{0}$, is just the average electron density n, and the other coefficients represent fluctuations about this average density. Equation (6.3) may be written in the form

$$\dot{\mathbf{v}}_i = -\frac{4\pi e^2 i}{m} \sum_{\mathbf{k}}{}' \frac{\mathbf{k}}{k^2} \rho_{\mathbf{k}} \exp(i\mathbf{k}.\mathbf{r}_i), \qquad \ldots\ldots(6.7)$$

which makes it clear why $\mathbf{k} = \mathbf{0}$ is excluded: the effect of $\rho_{\mathbf{0}}$ is cancelled by the uniform background of positive charge and so $\rho_{\mathbf{0}}$ cannot appear in the equation of motion.

From equation (6.6) we have

$$\dot{\rho}_{\mathbf{k}} = -i \sum_i (\mathbf{k}.\mathbf{v}_i) \exp(-i\mathbf{k}.\mathbf{r}_i), \qquad \ldots\ldots(6.8)$$

and
$$\ddot{\rho}_{\mathbf{k}} = -\sum_i [(\mathbf{k}.\mathbf{v}_i)^2 + i\mathbf{k}.\dot{\mathbf{v}}_i] \exp(-i\mathbf{k}.\mathbf{r}_i). \qquad \ldots\ldots(6.9)$$

Substitution of $\dot{\mathbf{v}}_i$ from equation (6.7) gives

$$\ddot{\rho}_{\mathbf{k}} = -\sum_i (\mathbf{k}.\mathbf{v}_i)^2 \exp(-i\mathbf{k}.\mathbf{r}_i) - \frac{4\pi e^2}{m}\sum_i \sum_{\mathbf{k}'}{}' \frac{\mathbf{k}.\mathbf{k}'}{k'^2}\rho_{\mathbf{k}'} \exp[i(\mathbf{k}'-\mathbf{k}).\mathbf{r}_i]. \qquad \ldots\ldots(6.10)$$

The last term may be split into two parts, consisting of those terms of the double summation with $\mathbf{k}'=\mathbf{k}$ and those with $\mathbf{k}'\neq\mathbf{k}$ respectively, as follows:

$$-\frac{4\pi e^2 n}{m}\rho_{\mathbf{k}} - \frac{4\pi e^2}{m}\sum_{\mathbf{k}'\neq\mathbf{k}}{}' \frac{\mathbf{k}.\mathbf{k}'}{k'^2}\rho_{\mathbf{k}'}\rho_{\mathbf{k}-\mathbf{k}'}. \qquad \ldots\ldots(6.11)$$

Now we may expect the Fourier coefficients $\rho_{\mathbf{k}}, \mathbf{k}\neq\mathbf{0}$, to be very small, since there is a large number of electrons and they are more or less randomly situated, so that in summing over i the phase factors $\exp(-i\mathbf{k}.\mathbf{r}_i)$ will tend to cancel out to zero. It is reasonable to suggest, therefore, that the second term in (6.11) will be small compared with the first, at least for small values of k. In fact, Bohm and Pines dispense with the second term entirely, and call this the ***random phase approximation***, which we shall see is also a central feature of their quantum-mechanical theory. Of course, this step requires more detailed consideration than it is possible to give here, and the reader is referred to the papers of Bohm and Pines (1952, 1953).

Using the random phase approximation, then, equation (6.10) becomes

$$\ddot{\rho}_{\mathbf{k}} = -\sum_i (\mathbf{k}.\mathbf{v}_i)^2 \exp(-i\mathbf{k}.\mathbf{r}_i) - \omega_{\mathrm{D}}^2\rho_{\mathbf{k}}. \qquad \ldots\ldots(6.12)$$

The first term on the right does not involve the accelerations and would be present even if the electrons did not interact—it arises simply from the thermal motion of the electrons. The second term represents the effect of the electronic interactions. It is clear that, for sufficiently small values of k, the first term, which contains the factor k^2, may be neglected in comparison with the second, giving

$$\ddot{\rho}_{\mathbf{k}} + \omega_{\mathrm{D}}^2\rho_{\mathbf{k}} = 0. \qquad \ldots\ldots(6.13)$$

Thus we see that the Fourier coefficients of the density, for sufficiently long wavelengths, oscillate with the plasma frequency; which is merely another expression of the result obtained in § 3. However, equation (6.12), which may be written

$$\ddot{\rho}_{\mathbf{k}} = -\sum_i [(\mathbf{k}.\mathbf{v}_i)^2 + \omega_{\mathrm{D}}^2] \exp(-i\mathbf{k}.\mathbf{r}_i), \qquad \ldots\ldots(6.14)$$

shows explicitly the effect of the random thermal motion of the electrons, which tends to disrupt the organized motion of the electron gas. We may say, very roughly, that organized oscillation is possible provided that

$$\omega_{\mathrm{D}}^2 \gg \langle(\mathbf{k}.\mathbf{v}_i)^2\rangle_{\mathrm{av}}. \qquad \ldots\ldots(6.15)$$

For a classical plasma with a Maxwellian velocity distribution, this becomes

$$k^2 \ll (3\omega_{\mathrm{D}}^2/\langle v_i^2\rangle_{\mathrm{av}}) = \lambda_{\mathrm{D}}^{-2}, \qquad \ldots\ldots(6.16)$$

where λ_{D} is the 'Debye length' (Debye and Hückel 1923). For a typical gas discharge λ_{D} is of the order of 10^{-3} cm, while for the $\langle v_i^2\rangle_{\mathrm{av}}$ occurring in a metal it is of the order of the interatomic spacing. The foregoing classical treatment can be extended to obtain the dispersion relation, equation (3.9), and also to obtain the effective screening of the charge of an electron, the screening distance again being found to be roughly λ_D. However, we shall not pursue this further but shall proceed with the quantum-mechanical theory appropriate to a metal.

§ 7. QUANTUM-MECHANICAL THEORY

We shall still restrict ourselves to the model of free electrons in a box of unit volume described in the previous section, but for the quantum-mechanical treatment we shall use a Hamiltonian formalism. It is sufficient to consider the Hamiltonian

$$H_0 = \sum_i \frac{p_i^2}{2m} + \tfrac{1}{2} \sum_{i,j \neq i} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \text{constant}. \qquad \ldots\ldots(7.1)$$

The first term represents the kinetic energy of the electrons and the second the potential energy due to their Coulomb interaction. The third term, which we have merely designated ' constant ', accounts for the uniform background of positive charge. In a real metal this term would contain the ion–ion interactions and the potential energy of the electrons in the field of the ions ; it would have to be considered explicitly in calculating the cohesive energy of a metal.

Using the Fourier expansion (6.1) the Hamiltonian becomes

$$H_0 = \sum_i \frac{p_i^2}{2m} + 2\pi e^2 \sum_{i,j \neq i} \sum_{\mathbf{k}}{}' \frac{1}{k^2} \exp\left[i\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)\right]. \qquad \ldots\ldots(7.2)$$

In the following work it will be understood that, in summing over $\mathbf{k}$, $\mathbf{k} = \mathbf{0}$ is to be excluded, and the prime on the summation sign will be omitted : this takes care of the background of positive charge, as explained in § 6, and cancels the constant term in equation (7.1).

The transition to quantum mechanics is made in the usual way by regarding the components of $\mathbf{p}_i$, the momentum of the ith electron, and the corresponding components of $\mathbf{r}_i$, for all i, as non-commuting operators. The eigenvalues E of the Hamiltonian are then given by the equation

$$H_0 \Psi = E \Psi, \qquad \ldots\ldots(7.3)$$

the wave function Ψ being a function of the coordinates, both space and spin, of all the electrons.

There are several ways of introducing the plasma oscillations. In their original work, for instance, Bohm and Pines (1953) started by re-writing the Hamiltonian (7.2) in terms of the vector potential of the longitudinal part of the electromagnetic field of the electrons ; we are, of course, consistently ignoring the transverse part, which gives the magnetic interactions. We shall, however, present a more formal treatment suggested by Pines (1954, 1955) which has the advantage of conciseness and possibly of showing more clearly the difficulties of the theory and the approximations involved in it.

The classical theory of plasma oscillations, both in its elementary and more sophisticated forms, leads us to expect that these oscillations will occur in metals, but only with wave numbers less than some k_c, of the order of 1 Å^{-1}. We might consequently anticipate the possibility of transforming the Hamiltonian (7.2) in such a way that the plasma oscillations show themselves explicitly as a finite set of simple harmonic oscillator Hamiltonians, one for each $\mathbf{k}$ with $k<k_c$, i.e. Hamiltonians of the type

$$\tfrac{1}{2}(P_{\mathbf{k}}{}^{*}P_{\mathbf{k}}+\omega^2 Q_{\mathbf{k}}{}^{*}Q_{\mathbf{k}}), \qquad \ldots\ldots(7.4)$$

where $Q_{\mathbf{k}}$ and $P_{\mathbf{k}}$ are the canonically conjugate coordinate and momentum relating to the oscillator. The set of $Q_{\mathbf{k}}$ and $P_{\mathbf{k}}$, in fact, describe part of the longitudinal electromagnetic field of the electrons; in other words, they are the Fourier coefficients in the expansions of real field quantities, such as the vector potential, of the type

$$Q(\mathbf{r})=\sum_{\mathbf{k}} Q_{\mathbf{k}}\hat{\mathbf{k}}\exp(i\mathbf{k}.\mathbf{r}), \qquad \ldots\ldots(7.5)$$

where $\hat{\mathbf{k}}$ denotes a unit vector in the direction of $\mathbf{k}$. To ensure that the field is real, therefore, we must impose the conditions

$$Q_{\mathbf{k}}{}^{*}=-Q_{-\mathbf{k}},\; P_{\mathbf{k}}{}^{*}=-P_{-\mathbf{k}}. \qquad \ldots\ldots(7.6)$$

Let us, therefore, quite formally introduce a set of $P_{\mathbf{k}}$, and implicitly a set of $Q_{\mathbf{k}}$, with $k<k_c$, by adding on to the Hamiltonian H_0 a term

$$H_1=\sum_{\mathbf{k}(k<k_c)}\left[\tfrac{1}{2}P_{\mathbf{k}}{}^{*}P_{\mathbf{k}}-i\left(\frac{4\pi e^2}{k^2}\right)^{1/2}P_{\mathbf{k}}\rho_{\mathbf{k}}\right], \qquad \ldots\ldots(7.7)$$

where $\rho_{\mathbf{k}}$ is defined by equation (6.6). This term is the sum of a field energy and a particle–field interaction energy, but the interpretation need not concern us—we may regard the addition of this term as a formal step to be justified by results. Of course, it is not permissible to take this step unconditionally, for in so doing we have increased the number of degrees of freedom of the system by $k_c{}^3/6\pi^2$, the number of the $Q_{\mathbf{k}}$, i.e. the number of $\mathbf{k}$-values in a sphere of radius k_c, the density being $1/8\pi^3$. Our aim, however, is merely to re-describe part of the electronic motion in terms of plasma oscillations, the number of degrees of freedom remaining fixed at $3n$. We therefore impose on our wave function Ψ the *subsidiary conditions*

$$P_{\mathbf{k}}\Psi=0 \quad \text{for} \quad k<k_c. \qquad \ldots\ldots(7.8)$$

According to the usual quantum-mechanical procedure it is assumed that the field coordinates and momenta satisfy the commutation relations

$$[P_{\mathbf{k}},Q_{\mathbf{k}'}]=P_{\mathbf{k}}Q_{\mathbf{k}'}-Q_{\mathbf{k}'}P_{\mathbf{k}}=-i\hbar\delta_{\mathbf{k}\mathbf{k}'}, \qquad \ldots\ldots(7.9)$$

so that $P_{\mathbf{k}}$ can be represented by the differential operator $-i\hbar\partial/\partial Q_{\mathbf{k}}$. In this representation, (7.8) becomes

$$\partial\Psi/\partial Q_{\mathbf{k}}=0 \quad \text{for} \quad k<k_c. \qquad \ldots\ldots(7.10)$$

Thus, the subsidiary conditions (7.8) merely ensure that Ψ does not contain any of the $Q_{\mathbf{k}}$, that is to say, Ψ must be a function of the electronic coordinates only. It is clear that for any wave function of this type the additional term in the Hamiltonian cannot affect the eigenvalues, for H_1 operates on Ψ to give zero; thus

$$(H_0+H_1)\Psi=H_0\Psi=E\Psi, \qquad \ldots\ldots(7.11)$$

with the same E as in (7.3).

It seems, therefore, that the method we have adopted so far is correct but it is also pointless, for it gives us no new information. To remedy this we shall make a canonical transformation, that is, a transformation of operators which leaves the eigenvalues of the Hamiltonian unaltered, and this will be deliberately chosen to give results of the form suggested by the classical theory. The test of this procedure will be in the possibility of then choosing k_c, which at present is arbitrary, so that these results are physically significant. The required transformation is that under which an operator A becomes $S^{-1}AS$, where S is the unitary operator

$$S = \exp\left[\frac{1}{\hbar}\sum_{\mathbf{k}(k<k_c)}\left(\frac{4\pi e^2}{k^2}\right)^{1/2} Q_{\mathbf{k}}\rho_{\mathbf{k}}^*\right]. \quad \ldots\ldots(7.12)$$

It is easily verified, using equation (7.6), that $S^*S = S^{-1}S = 1$.

The various operators transform in the following way:

$$\left.\begin{aligned} &\mathbf{r}_i \to S^{-1}\mathbf{r}_i\, S = \mathbf{r}_i, \quad Q_{\mathbf{k}} \to S^{-1} Q_{\mathbf{k}} S = Q_{\mathbf{k}}, \\ &\mathbf{p}_i \to S^{-1}\mathbf{p}_i\, S = \mathbf{p}_i + (4\pi e^2)^{1/2} \sum_{\mathbf{k}(k<k_c)} Q_{\mathbf{k}}\hat{\mathbf{k}} \exp(i\mathbf{k}.\mathbf{r}_i), \\ &P_{\mathbf{k}} \to S^{-1}P_{\mathbf{k}}S = P_{\mathbf{k}} - i\left(\frac{4\pi e^2}{k^2}\right)^{1/2}\rho_{\mathbf{k}}^*. \end{aligned}\right\} \quad \ldots\ldots(7.13)$$

As an illustration we shall perform the transformation in detail for $P_{\mathbf{k}}$. We have

$$\begin{aligned} S^{-1}P_{\mathbf{k}}S &= S^{-1}SP_{\mathbf{k}} + S^{-1}P_{\mathbf{k}}S - S^{-1}SP_{\mathbf{k}} \\ &= P_{\mathbf{k}} + S^{-1}[P_{\mathbf{k}}, S] \\ &= P_{\mathbf{k}} - i\hbar S^{-1}\partial S/\partial Q_{\mathbf{k}} \\ &= P_{\mathbf{k}} - i\left(\frac{4\pi e^2}{k^2}\right)^{1/2}\rho_{\mathbf{k}}^*. \end{aligned} \quad \ldots\ldots(7.14)$$

The same procedure applies to $Q_{\mathbf{k}}$ and to each component of $\mathbf{p}_i$ and $\mathbf{r}_i$.

The subsidiary conditions therefore become

$$S^{-1}P_{\mathbf{k}}SS^{-1}\Psi = 0 \quad \text{for} \quad k < k_c,$$

or

$$\left[P_{\mathbf{k}} - i\left(\frac{4\pi e^2}{k^2}\right)^{1/2}\rho_{\mathbf{k}}^*\right]\Phi = 0 \quad \text{for} \quad k < k_c, \quad \ldots\ldots(7.15)$$

where $\Phi = S^{-1}\Psi$ is the new wave function. This equation is significant because it shows that the canonical transformation has related the field momenta $P_{\mathbf{k}}$ to the density fluctuations $\rho_{\mathbf{k}}$ for long wavelengths, and this is very promising, for the classical theory shows that the latter oscillate with roughly the plasma frequency.

The Hamiltonian $H = H_0 + H_1$ transforms in the following way:

$$H \to S^{-1}HS = \mathscr{H},$$

where

$$\begin{aligned} \mathscr{H} = {} & \sum_i \frac{p_i^2}{2m} + \frac{1}{2}\sum_{\mathbf{k}(k<k_c)} (P_{\mathbf{k}}^* P_{\mathbf{k}} + \omega_p^2 Q_{\mathbf{k}}^* Q_{\mathbf{k}}) \\ & + 2\pi e^2 \sum_{i, j\neq i}\sum_{\mathbf{k}(k>k_c)} \frac{1}{k^2}\exp[i\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)] - 2\pi e^2 n \sum_{\mathbf{k}(k<k_c)} \frac{1}{k^2} \\ & + \left(\frac{4\pi e^2}{m^2}\right)^{1/2} \sum_i \sum_{\mathbf{k}(k<k_c)} \hat{\mathbf{k}}.(\mathbf{p}_i - \tfrac{1}{2}\hbar\mathbf{k})\, Q_{\mathbf{k}} \exp(i\mathbf{k}.\mathbf{r}_i) \\ & + \frac{2\pi e^2}{m} \sum_i \sum_{\substack{\mathbf{k}, \mathbf{l} \neq -\mathbf{k} \\ (k, l < k_c)}} \hat{\mathbf{k}}.\hat{\mathbf{l}}\, Q_{\mathbf{k}} Q_{\mathbf{l}} \exp[i(\mathbf{k}+\mathbf{l}).\mathbf{r}_i]. \end{aligned} \quad \ldots\ldots(7.16)$$

This transformation is easily accomplished using equations (7.6) and (7.13) and remembering that $\mathbf{p}_i$ and $\mathbf{r}_i$ satisfy the usual commutation relations. We wish only to point out that, as in the classical case, equation (6.11), one sum has been split into two parts, with $\mathbf{k} = -\mathbf{l}$ and with $\mathbf{k} \neq -\mathbf{l}$ respectively; the former gives

$$\tfrac{1}{2} \sum_{\mathbf{k}(k<k_c)} \omega_p^2 Q_{\mathbf{k}}^* Q_{\mathbf{k}},$$

with $\omega_p^2 = 4\pi e^2 n/m$ as usual, and the latter is the final term in $\mathscr{H}$. The eigenvalue equation (7.11) becomes

$$S^{-1}HSS^{-1}\Psi = ES^{-1}\Psi,$$

or

$$\mathscr{H}\Phi = E\Phi, \qquad \ldots\ldots(7.17)$$

with eigenvalues E unchanged. In what follows we shall be concerned exclusively with the ground state of the system, so in order to avoid complexity of notation we shall let E and Φ denote the energy and wave function respectively of the ground state.

Let us now discuss the various terms appearing in $\mathscr{H}$. According to the *random phase approximation*, introduced in the classical treatment, the last term may be neglected. This step is justified by Bohm and Pines (1953, p. 614). In the same place these authors show that the effect of the second last term, which represents an electron–plasma interaction, is small for the value of k_c eventually chosen, and for the present we shall neglect this term also. This leaves us with the first four terms only.

The first term is just the particle kinetic energy. The second term is a sum of harmonic oscillator Hamiltonians and represents the plasma oscillations, each having angular frequency ω_p. The third term is the residual screened Coulomb interaction $H_{\text{s.r.}}$, which has already been discussed in § 5; if the sum over $\mathbf{k}$ is replaced by an integral in the usual way, the density of representative points in $\mathbf{k}$-space being $1/8\pi^3$, we obtain

$$\begin{aligned} H_{\text{s.r.}} &= \frac{2\pi e^2}{8\pi^3} \sum_{i,j\neq i} \int_{(k>k_c)} \frac{1}{k^2} \exp[i\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)]\, d\mathbf{k} \\ &= \tfrac{1}{2} \sum_{i,j\neq i} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} F(k_c|\mathbf{r}_i - \mathbf{r}_j|), \qquad \ldots\ldots(7.18) \end{aligned}$$

where the function F is given by equation (5.4) and its graph is shown in figure 2.

In order to explain the presence of the fourth term we shall formally split up the total Coulomb interaction, the second term in equation (7.2), in the following way:

$$\begin{aligned} &2\pi e^2 \sum_{i,j\neq i} \sum_{\mathbf{k}} \frac{1}{k^2} \exp[i\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)] \\ &= H_{\text{s.r.}} + 2\pi e^2 \sum_{i,j} \sum_{\mathbf{k}(k<k_c)} \frac{1}{k^2} \exp[i\mathbf{k}.(\mathbf{r}_i - \mathbf{r}_j)] - 2\pi e^2 n \sum_{\mathbf{k}(k<k_c)} \frac{1}{k^2}. \qquad \ldots\ldots(7.19) \end{aligned}$$

During the transformation of H the second term on the right-hand side, in which it is not stipulated that $j \neq i$, is automatically absorbed in the plasma part of $\mathscr{H}$, leaving the other two terms to appear explicitly.

So long as the last two terms of $\mathscr{H}$ are negligible, therefore, we have achieved our aim of re-describing the system as a set of plasma oscillations plus a set of particles which have only a short-range interaction. It remains to calculate k_c; but first we must discuss the wave function Φ.

It is necessary to point out here that the function Φ, as well as being a solution of the eigenvalue equation (7.17), must also satisfy the subsidiary conditions (7.15). Bohm and Pines (1953, p. 613) have given an argument by which they claim to prove that the *ground state* wave function automatically satisfies the subsidiary conditions, but, using a different approach, Adams (1955 b) has raised serious doubts as to the possibility of ever finding a proper wave function which would satisfy the subsidiary conditions exactly in their present form. It would be out of place to dwell upon this problem at length in an account of the present kind, particularly as no satisfactory solution of it is forthcoming (see, however, Pines 1955, p. 390). Certainly a difficulty exists and the quantum-mechanical theory described here must be viewed with reserve until it is cleared up; however, the results of the theory are so desirable and the physical picture it presents is so reasonable that one is encouraged to believe that this difficulty is not a serious one. Similar difficulties have occurred in the quantum theory of radiation and in nuclear physics without being fatal. As far as detailed calculation goes we shall be concerned only with the ground state of the system and will ignore the subsidiary conditions.

The function Φ is a function of the field coordinates $Q_{\mathbf{k}}$ and the particle coordinates $\mathbf{r}_i$, but if the electron–plasma interaction is negligible, as we have assumed, Φ may be expressed as the product of a function of the $Q_{\mathbf{k}}$ and a function of the $\mathbf{r}_i$. If indeed k_c^{-1} is only about an ångström unit, and eventual calculation will confirm this, the residual screened interaction is of such short range that a good initial approximation to the wave function is obtained by treating the particles as if they were completely free, i.e. non-interacting. In this case the ground state wave function can be written in the approximate form

$$\Phi = G(\ldots Q_{\mathbf{k}} \ldots)\, D_0(\ldots \mathbf{r}_i' \ldots), \qquad \ldots\ldots(7.20)$$

where G is a product of harmonic oscillator ground state wave functions, one for each $Q_{\mathbf{k}}$, and D_0 is the determinantal function, equation (2.8), describing the ground state of a system of n free electrons.

The energy corresponding to this wave function is just the sum of the zero-point energies of the oscillators, $\frac{1}{2}\hbar\omega_p$ for each, and the energy of the particles; thus

$$E = \tfrac{3}{5}n\zeta_0 + \sum_{\mathbf{k}(k<k_c)} \left(\frac{\hbar\omega_p}{2} - \frac{2\pi n e^2}{k^2}\right) + \langle H_{\mathrm{s.r.}}\rangle_{\mathrm{av}}. \qquad \ldots\ldots(7.21)$$

The first term is the kinetic energy of the particles. The second term, if the sum is replaced by an integral as described before, becomes

$$\frac{\hbar\omega_p k_c^3}{12\pi^2} - \frac{n e^2 k_c}{\pi}. \qquad \ldots\ldots(7.22)$$

If we introduce the dimensionless quantity

$$\beta = k_c/k_0, \qquad \ldots\ldots(7.23)$$

and use the fact that $k_0^3 = 3\pi^2 n$, (7.22) becomes

$$\frac{n\hbar\omega_{\text{p}}}{4}\beta^3 - \frac{ne^2k_0}{\pi}\beta. \qquad \ldots\ldots(7.24)$$

The last term in (7.21) is the exchange energy due to the use of a determinantal particle wave function. It is smaller than the exchange energy in the Hartree–Fock theory because here the interaction term is $H_{\text{s.r.}}$ instead of the whole Coulomb interaction. Its calculation, although straightforward, is rather lengthy and will not be reproduced here; we merely quote the result obtained by Pines (1953, Appendix 1):

$$\langle H_{\text{s.r.}}\rangle_{\text{av}} = -\frac{3ne^2k_0}{4\pi}\left(1-\frac{4}{3}\beta+\frac{\beta^2}{2}-\frac{\beta^4}{48}\right). \qquad \ldots\ldots(7.25)$$

Thus, the total energy per electron is

$$\begin{aligned}\frac{E}{n} &= \frac{3}{5}\zeta_0+\frac{\hbar\omega_{\text{p}}}{4}\beta^3-\frac{e^2k_0}{\pi}\beta \\ &\quad -\frac{3e^2k_0}{4\pi}\left(1-\frac{4}{3}\beta+\frac{\beta^2}{2}-\frac{\beta^4}{48}\right) \\ &= \frac{3}{5}\zeta_0+\frac{\hbar\omega_{\text{p}}}{4}\beta^3-\frac{3e^2k_0}{4\pi}\left(1+\frac{\beta^2}{2}-\frac{\beta^4}{48}\right). \qquad \ldots\ldots(7.26)\end{aligned}$$

If we now express energies in rydbergs and lengths in Bohr units, and use formula (4.2) for $\hbar\omega_{\text{p}}$, we obtain

$$\begin{aligned}\frac{E}{n} &= \frac{2\cdot21}{r_{\text{e}}^2}+\frac{\sqrt{3}}{2r_{\text{e}}^{3/2}}\beta^3-\frac{0\cdot916}{r_{\text{e}}}\left(1+\frac{\beta^2}{2}-\frac{\beta^4}{48}\right) \\ &= \frac{2\cdot21}{r_{\text{e}}^2}-\frac{0\cdot916}{r_{\text{e}}}+\frac{0\cdot866}{r_{\text{e}}^{3/2}}\beta^3-\frac{0\cdot458}{r_{\text{e}}}\beta^2+\frac{0\cdot019}{r_{\text{e}}}\beta^4 \text{ ryd.} \qquad \ldots\ldots(7.27)\end{aligned}$$

Now, although we have had to make certain assumptions about its order of magnitude, k_{c} is at the moment virtually an arbitrary parameter which determines the number of oscillator functions appearing in the wave function (7.20). If this were, in fact, the whole story, a perfectly correct method of determining β, and hence k_{c}, would be to minimize the above expression for the energy with respect to β. However, k_{c} also determines the number of field variables introduced into the Hamiltonian and the number of subsidiary conditions, so that the latter should really be taken into account in the variational procedure. This point has been discussed by Pines (1955, p. 363) and difficulties which might arise have been demonstrated by Kuper (1956). None the less, for want of a better method which is also tractable, we shall follow Pines in ignoring the subsidiary conditions and minimize (7.27) directly with respect to β. If we drop the small term in β^4, which is in fact of the same order as terms which have been neglected in the approximations made so far, we find

$$\frac{\partial}{\partial\beta}\left(\frac{E}{n}\right) = \frac{2\cdot598}{r_{\text{e}}^{3/2}}\beta^2-\frac{0\cdot916}{r_{\text{e}}}\beta. \qquad \ldots\ldots(7.28)$$

Equating this to zero we obtain

$$\beta = 0\cdot353r_{\text{e}}^{1/2}, \qquad \ldots\ldots(7.29)$$

or

$$k_{\text{c}} = 0\cdot353r_{\text{e}}^{1/2}k_0 = 0\cdot677r_{\text{e}}^{-1/2}. \qquad \ldots\ldots(7.30)$$

It must be remembered that here r_e is in Bohr units. Taking $r_e = 4$ B.u., for *sodium* we obtain $k_c^{-1} = 2{\cdot}95$ B.u. $= 1{\cdot}56$ Å. The screening distance k_c^{-1} is not much greater than this for any metal, and for most of them it is smaller, so that our assumption that this was small is well justified.

The above procedure of minimizing E is equivalent to minimizing the *long-range correlation energy*, a quantity which will be discussed below. Of course, although (7.27) is the total energy corresponding to the wave function (7.20), it is only an approximation to the true ground state energy, since (7.20) would be of the correct form only if the particles did not interact. The particle part of the wave function should, in fact, be an infinite sum of determinants of free-electron functions, corresponding to the ground state and excited states of a system of free electrons, with coefficients determined by $H_{\text{s.r.}}$. Fortunately, since the screened interaction is of such short range, the effect of $H_{\text{s.r.}}$ can be found by second-order perturbation theory, as will be shown in the following section. The energy correction is called the *short-range correlation energy*.

In all the foregoing work it has been assumed that the last two terms of $\mathscr{H}$, equation (7.16), are negligible. This is certainly true of the last term but not quite true of the second last term, the electron–plasma interaction. However, by means of a further canonical transformation Bohm and Pines (1953) have been able to eliminate almost completely the electron–plasma interaction. During this process some of the other terms are altered, but not drastically, and certain very small new terms appear. Of the four principal terms in $\mathscr{H}$ only the first two are altered, and these become

$$\sum_i \frac{p_i^2}{2m}\left(1-\frac{\beta^3}{6}\right)+\tfrac{1}{2}\sum_{\mathbf{k}(k<k_c)}(P_{\mathbf{k}}^{*}P_{\mathbf{k}}+\omega^2 Q_{\mathbf{k}}^{*}Q_{\mathbf{k}}), \qquad \ldots\ldots(7.31)$$

where

$$\omega \simeq \omega_p\left(1+\frac{3\zeta_0 k^2}{5m\omega_p^2}+\frac{\hbar^2 k^4}{8m^2\omega_p^2}\right). \qquad \ldots\ldots(7.32)$$

Thus we see that the essential effects of the second canonical transformation are to change the mass of the particles from m to an ' effective ' mass $m/(1-\beta^3/6)$, and to make the frequencies of the modes of plasma oscillation depend upon the wave number k. Since $\beta < 1$, the effective mass does not differ much from m, but this difference emphasizes the fact that the particles whose interaction is $H_{\text{s.r.}}$ are not electrons, but electrons associated in some way with the plasma oscillations. If both sides of (7.32) are squared and terms containing orders of k higher than the second then omitted, we get the classical dispersion relation, equation (3.9), since

$$\tfrac{3}{5}\zeta_0 = \tfrac{1}{2}m\langle v_i^2\rangle_{\text{av}}. \qquad \ldots\ldots(7.33)$$

§ 8. THE CORRELATION ENERGY

8.1. *The Long-range Correlation Energy*

The correlation energy has been defined in § 2 as the difference between the energy calculated by the Hartree–Fock method and the energy obtained when the Coulomb correlations are correctly taken into account. The Hartree–Fock energy per electron is

$$E_{\text{HF}} = \frac{2{\cdot}21}{r_e^2} - \frac{0{\cdot}916}{r_e}\ \text{ryd.} \qquad \ldots\ldots(8.1)$$

The energy (7.27), in which the second-order perturbing effect of $H_{\text{s.r.}}$ is not included, is already lower than (8.1), and the difference is called the ***long-range correlation energy***, which we shall denote by W_l. Again neglecting the term in β^4, we find

$$W_l = \frac{0{\cdot}866}{r_e^{3/2}}\beta^3 - \frac{0{\cdot}458}{r_e}\beta^2 \text{ ryd.} \qquad \ldots\ldots(8.2)$$

The reason for this difference, and for the name, is simply that, owing to the factor G in the wave function (7.20), long-range Coulomb correlations are more adequately treated than in the Hartree–Fock theory. The use of a determinantal wave function in the latter has, as explained in § 2, the effect of reducing the probability of two electrons with parallel spins approaching one another closely, but the electronic motions are otherwise uncorrelated. In the plasma theory leading to (7.27), however, long-range Coulomb correlations irrespective of the spins of the electrons are included in the plasma oscillations, and it is only thereafter that the Pauli principle implicit in the determinantal particle function comes into action. As well as lowering the energy this has the result of eliminating the exaggerated 'exchange' effects associated with the Hartree–Fock theory, as will be seen in § 9.

Taking $\beta = 0{\cdot}353 r_e^{1/2}$, as found in § 7, (8.2) becomes

$$W_l = -0{\cdot}019 + 0{\cdot}0003 r_e \text{ ryd.} \qquad \ldots\ldots(8.3)$$

The small correction to this due to the electron–plasma interaction is given by Pines (1955) as $-0{\cdot}004$ ryd, so that the final long-range correlation energy per electron is

$$W_l = -0{\cdot}023 + 0{\cdot}0003 r_e \text{ ryd.} \qquad \ldots\ldots(8.4)$$

8.2. *The Short-range Correlation Energy*

The term $\langle H_{\text{s.r.}}\rangle_{\text{av}}$ in equation (7.21) is the first-order perturbation correction, obtained by using a single determinantal particle wave function which would only be correct if no interaction term $H_{\text{s.r.}}$ appeared. We must now obtain the energy correction due to the perturbing effect of $H_{\text{s.r.}}$ on the wave function, i.e. to the use of a particle wave function consisting of an infinite sum of determinants of free-electron functions instead of the single determinant D_0 in (7.20). Since it is the result of short-range correlations among the electronic motions this energy correction is called the ***short-range correlation energy***, and we shall denote it ***per electron*** by W_s. Most of the necessary algebra is given in the paper by Macke (1950) mentioned in § 2. Macke, however, treated the whole Coulomb interaction as a perturbation, which causes the second-order correction to diverge. Here, because $H_{\text{s.r.}}$ is so weak, the second-order perturbation energy correction is quite small and we shall assume that it gives W_s with sufficient accuracy. For a further discussion of the validity of using perturbation theory the reader may consult the article by Pines (1955).

We have, then,

$$W_s = -\frac{1}{n}\sum_{j\neq 0}\frac{|(H_{\text{s.r.}})_{0j}|^2}{E_j - E_0}, \qquad \ldots\ldots(8.5)$$

where $(H_{\text{s.r.}})_{0j}$ is the matrix element of $H_{\text{s.r.}}$ between the determinant D_0 and a determinant D_j representing a virtual excited state, i.e. a determinant like D_0 in which some of the free-electron functions are those of excited one-electron states lying outside the Fermi sphere. E_0 and E_j are the corresponding kinetic energies, i.e. the energies of the system of particles in the states D_0 and D_j when no interaction is present. It is easy to show, in fact, that the only non-zero matrix elements are those for which D_j contains *two and only two* excited one-electron functions, and we shall only consider these.

In going from D_0 to D_j the two excited electrons can, of course, have either parallel or antiparallel spins. The matrix element is zero, however, unless the spins of the excited states are the same as those of the initial states. The terms in (8.5) corresponding to parallel spin excitations give the part of the short-range correlation energy due to parallel spin correlations and the remaining terms give the part due to antiparallel spin correlations. Now we know that the Pauli principle has the effect of keeping electrons with parallel spins apart, reducing the effectiveness of $H_{\mathrm{s.r.}}$, so it is reasonable to assume that the parallel spin part of W_s will be small. The likelihood of this is borne out by the mathematical expression (Pines 1955, eqn. (6.21)), which consists of two similar terms with opposite signs ; however, this expression has not yet been evaluated, so we cannot say with certainty that it is completely negligible. None the less, we shall assume that almost the whole of W_s is due to antiparallel spin correlations.

The only D_j we shall consider, then, are those obtained from D_0 by raising two electrons with wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$, and opposite spins, to excited levels with wave vectors $\mathbf{k}_1+\mathbf{k}$ and $\mathbf{k}_2-\mathbf{k}$ lying outside the Fermi sphere, the total momentum thus being conserved. We then find

$$(H_{\mathrm{s.r.}})_{0j} = 4\pi e^2/k^2, \qquad \ldots\ldots(8.6)$$

$$\begin{aligned} E_j - E_0 &= \frac{\hbar^2}{2m}[(\mathbf{k}_1+\mathbf{k})^2+(\mathbf{k}_2-\mathbf{k})^2-k_1{}^2-k_2{}^2] \\ &= \frac{\hbar^2}{m}\mathbf{k}\,.\,(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k}), \end{aligned} \qquad \ldots\ldots(8.7)$$

and

$$W_s = -\frac{1}{n}\sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}} \frac{m(4\pi e^2)^2}{\hbar^2 k^4 \mathbf{k}\,.\,(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k})}, \qquad \ldots\ldots(8.8)$$

where $k>k_c$, and $\mathbf{k}_1, \mathbf{k}_2$ lie inside the Fermi sphere while $\mathbf{k}_1+\mathbf{k}, \mathbf{k}_2-\mathbf{k}$ lie outside it.

If we now replace the summations by integrations, we obtain

$$W_s = -\frac{m(4\pi e^2)^2}{n\hbar^2(8\pi^3)^3}\iiint \frac{d\mathbf{k}\,d\mathbf{k}_1\,d\mathbf{k}_2}{k^4\mathbf{k}\,.\,(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k})}, \qquad \ldots\ldots(8.9)$$

the region of integration being

$$k_1, k_2 < k_0,\ k>k_c,\ |\mathbf{k}_1+\mathbf{k}|>k_0,\ |\mathbf{k}_2-\mathbf{k}|>k_0.$$

Expressing k, k_1 and k_2 in units of k_0 and the energy in rydbergs, we have finally

$$W_s = -\frac{3}{16\pi^5}\iiint \frac{d\mathbf{k}\,d\mathbf{k}_1\,d\mathbf{k}_2}{k^4\mathbf{k}\,.\,(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k})}\ \mathrm{ryd}, \qquad \ldots\ldots(8.10)$$

where $k_1, k_2 < 1$, $k>\beta$, $|\mathbf{k}_1+\mathbf{k}|>1$, $|\mathbf{k}_2-\mathbf{k}|>1$.

The integrals can be evaluated without too much difficulty, but the calculation is lengthy and we shall again merely quote the result given by Pines :

$$W_s = -(0{\cdot}0254 - 0{\cdot}0626\ln\beta + 0{\cdot}00637\beta^2)\ \mathrm{ryd}, \qquad \ldots\ldots(8.11)$$

or, using $\beta = 0{\cdot}353 r_e^{1/2}$,

$$W_s = -(0{\cdot}0905 - 0{\cdot}0313 \ln r_e + 0{\cdot}0008 r_e) \text{ ryd.} \qquad \ldots\ldots(8.12)$$

The total correlation energy W per electron, the sum of (8.4) and (8.12), is thus

$$W = -(0{\cdot}114 - 0{\cdot}0313 \ln r_e + 0{\cdot}0005 r_e) \text{ ryd.} \qquad \ldots\ldots(8.13)$$

This result is in good agreement with Wigner's formula, equation (2.13). For sodium, with $r_e = 4$ B.u., (8.13) gives $W = 0{\cdot}073$ ryd, while Wigner's formula gives $W = 0{\cdot}064$ ryd. Pines (1955, p. 375) has emended Wigner's formula, using a more correct low density limit, and finds even better agreement with (8.13).

The plasma theory thus gives a correction to the Hartree–Fock theory of the size which is needed in cohesive energy calculations. We shall now see how it also improves on the Hartree–Fock theory in calculations which depend upon the distribution of one-electron energy levels.

§ 9. THE EFFECT OF PLASMA OSCILLATIONS ON THE ENERGY BAND THEORY OF METALS

The difficulties associated with the Hartree–Fock density of states are due to the magnitude of the exchange energy, i.e. the second term in equation (2.10), and the way in which it varies with **k**. In the plasma theory, after the long-range Coulomb correlations have been described by the plasma oscillations, the exchange energy of the particles, which comes from $H_{\text{s.r.}}$, is much smaller. We have already seen that this is so in the case of the total exchange energy $\langle H_{\text{s.r.}}\rangle_{\text{av}}$, equation (7.25); the ratio of this to the Hartree–Fock total exchange energy is

$$1 - \frac{4}{3}\beta + \frac{\beta^2}{2} - \frac{\beta^4}{48}, \qquad \ldots\ldots(9.1)$$

which amounts to about 0·3 for sodium. Here, however, we are more interested in the exchange energy, $\epsilon_x(\mathbf{k})$, associated with wave vector **k**. This has been calculated by Pines (1953, Appendix 1) and is as follows:

for $k < k_0 - k_c$,

$$\epsilon_x(\mathbf{k}) = -\frac{e^2}{2\pi}\left[2k_0 + \frac{k_0^2 - k^2}{k}\ln\left(\frac{k_0 + k}{k_0 - k}\right) - 4k_c\right], \qquad \ldots\ldots(9.2a)$$

for $k_0 - k_c < k$,

$$\epsilon_x(\mathbf{k}) = -\frac{e^2}{2\pi}\left[k_0 + \frac{k_0^2 - k^2}{k}\ln\left(\frac{k_0 + k}{k_c}\right) + \frac{3k^2 + k_c^2 - k_0^2}{2k} - 2k_c\right]. \qquad \ldots\ldots(9.2b)$$

If these expressions are used instead of the exchange term in equation (2.10), and the density of states computed in the usual way, the result is a distribution of the type shown in curve (*b*) of figure 3 (cf. Raimes 1954 b). The Hartree distribution is also shown, curve (*a*), for comparison. We see that both the band width and the density of states near the Fermi surface are now very similar to those of the Hartree theory (and, of course, the Sommerfeld theory). Without detailed calculation, therefore, we may conclude that the difficulties regarding electronic specific heats, x-ray band widths, etc., described in § 2, have been largely overcome by the use of the plasma theory.

There is one obvious difference between curve (*a*) and curve (*b*) in figure 3. The latter, which we shall call the Bohm–Pines distribution curve, has a discontinuity of slope at the energy corresponding to $k = k_0 - k_c$. Describing this in another way, we may say that the Bohm–Pines curve is similar to the Hartree curve except that it has a small ' tail '. The very pronounced appearance of this tail is undoubtedly due to the assumption of a sharp cut-off for the plasma oscillations, which slightly oversimplifies the problem. It might be expected that further refinements in the theory would result in a smooth curve, but it seems unlikely that this would have the parabolic form of the Hartree theory. Since the usual interpretation of the soft X-ray emission spectra of metals (Skinner 1940) is based upon the latter, this departure from the parabolic form must affect the empirical values of the band widths deduced from such spectra ; a discussion of this is given by Raimes (1954 b, c).

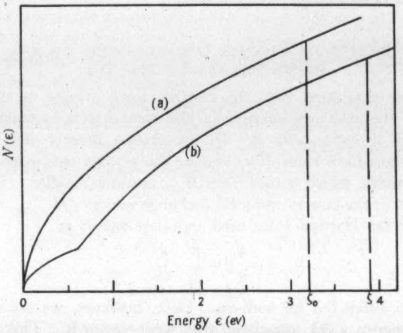


Figure 3. Density of states $N(\epsilon)$ for metallic sodium : (*a*) according to the Sommerfeld and Hartree theories ; (*b*) according to the theory of Bohm and Pines (all for free electrons). ζ_0 and ζ are the maximum energies at 0°K in the two cases.

In § 8 we have been concerned only with the *average* short-range correlation energy. An expression for the short-range correlation energy as a function of k is easily obtained, but it is difficult to evaluate except for $k = 0$ and $k = k_0$ (Pines 1955, Raimes 1954 b, c). The calculations for these extreme values of k suggest, however, that the effect of short-range correlations upon the band width and upon the $N(\epsilon)$ curve as a whole is very small.

Perhaps it should be mentioned that, in calculating the Bohm–Pines $N(\epsilon)$ curve shown in figure 3, the present author used the effective mass $m/(1-\beta^3/6)$, instead of the electronic mass m, in the kinetic energy term, a procedure which is suggested by the expression (7.31). In a private communication recently, however, Pines has suggested that it might be preferable to use just m, since the difference in energy due to the effective mass will probably be cancelled by the small additional terms in the Hamiltonian, after the second canonical transformation, which we have not considered in detail. The resulting difference in the band width is only a few per cent.

Further details of the beneficial effect of the plasma theory upon the calculation of various metallic properties may be found in the article by Pines (1955). However, these theoretical successes would prove somewhat unsatisfying were not more tangible evidence available of the existence of plasma oscillations in metals, and this we shall now discuss.

§ 10. ENERGY LOSSES OF FAST ELECTRONS IN PENETRATING METALLIC FILMS

It has been shown in § 4 that plasma oscillations are unlikely to be excited by heating, since even at quite high temperatures no electron has energy $\hbar\omega_p$ to give away. The utility of the plasma theory as a computational tool depends upon this fact. However, if a very fast electron enters a metal from outside there should be a good chance that it will excite one or more modes of plasma oscillation. If a beam of high energy electrons is shot through a thin metallic film, therefore, and we assume that the electrons in the metal are free, we should expect some of the emergent electrons to have lost energy $\hbar\omega_p$, others $2\hbar\omega_p$, others $3\hbar\omega_p$, and so on, according to the number of plasma quanta they have lost. Pines (1956) has given the name *plasmon* to such a quantum of plasma energy, and it will be convenient to use this name in the following account. Of course, even in a free-electron metal the losses will not be in multiples of exactly $\hbar\omega_p$ owing to dispersion, equation (7.32), but this is slight and we shall ignore it for the moment.

In fact, many experiments have now been carried out in which beams of electrons with energies of several kilovolts are passed through metallic films, generally of the order of 100 Å in thickness, and the energy loss spectra determined. Such high energies are not strictly required, since the excitation energy $\hbar\omega_p$ is in the region of 10 ev, but they are used for experimental convenience; the electrons are thus scattered through only a small angle in passing through the film. The first experiments of this type were by Ruthemann (1941, 1942, 1948) and Lang (1948); experimental details and results of these and more recent work are given in review articles by Marton, Leder and Mendlowitz (1955) and by Pines (1955, 1956). It is found that some metals, in particular beryllium, magnesium and aluminium, show several sharp loss peaks at almost exact integral multiples of $\hbar\omega_p$, the electron density used in computing ω_p being that of the *valence* electrons in the metal, i.e. two per atom for Be and Mg and three per atom for Al. The energy spectrum found by Blackstock, Ritchie and Birkhoff (1955) for aluminium, using a 45 kev beam, is shown in figure 4 as a striking example; $\hbar\omega_p$ for this metal is 16 ev and the experimental losses appear at intervals of very nearly 15 ev. However, many other metals, and curiously enough several non-metals, such as carbon, silicon, and germanium, show at least one pronounced loss peak at about $\hbar\omega_p$, where again the density of valence electrons, four per atom for C, Si and Ge, is used in computing ω_p.

On the other hand, agreement between the free-electron plasma theory and the experimental loss spectra is completely lacking for metals such as copper, silver, gold and the transition metals with well-filled d shells. Copper, for instance, shows a single broad peak at about $2\hbar\omega_p$, when the electron density is taken to be one per atom. This is not very surprising, for the free-electron theory is hardly likely to be valid where the ion cores are large or where the valence electron wave functions do not approximate closely to free-electron functions. However, this same argument should apply to carbon, silicon and germanium, and it does not appear to do so; a possible reason for this will be given in the next section.

Energy loss spectra are also disappointing in the case of the alkali metals, for which we expect good results from the free-electron theory. However, the alkali metals are particularly difficult from an experimental point of view, owing to the ease with which they oxidize, so that the results of energy loss experiments may not be reliable. We shall see in the following section that optical data provide experimental evidence of another kind which confirms the applicability of the free-electron plasma theory to these metals.

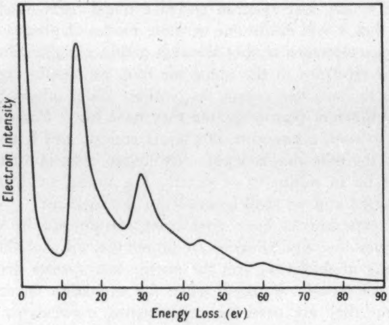


Figure 4. Energy loss spectrum of 45 kev electrons after passing through an aluminium film of thickness about 550 Å. Peaks occur at intervals of approximately 15 ev, while $\hbar\omega_p$ for this metal is 16 ev (*after* Blackstock, Ritchie and Birkhoff 1955).

Of course, the possibility suggests itself that the discrete energy losses may not be due to the emission of plasmons at all, but to the excitation of single electrons to higher energy zones. It is very likely that such single electron excitations occur, since the energy loss spectra of most metals show a number of small peaks which can hardly be due to excitation of collective oscillations (cf. Ferrell 1956). However, there is some experimental evidence, due to Watanabe (1956), that the well-defined loss peaks in Be, Mg, Al and Ge, at least, are the result of plasmon emission. Watanabe has measured the energy losses as a function of the angle through which the electrons are scattered, and obtained results which agree very well with the dispersion relation of Bohm and Pines, see equation (7.32) above.

Let us restrict ourselves to the case of a single plasmon emission of energy $\hbar\omega$. The relation between the momentum $\mathbf{p}$ of a fast incoming electron, the momentum $\hbar\mathbf{k}$ of the emitted plasmon, and the small angle θ through which the electron is scattered is shown in figure 5. Also, let us denote by U the initial energy of the electron, and by Δp the very small decrease in the magnitude of its momentum, i.e. $|\mathbf{p}-\hbar\mathbf{k}| = p-\Delta p$. In the experiments of Watanabe, U was 25 000 ev, $\hbar\omega$ in the region of 20 ev, and the maximum value of θ was about 10^{-2} radian. Now it is easily shown, according to the conservation of energy, that

$$\frac{\hbar\omega}{2U} = \frac{\Delta p}{p} - \frac{1}{2}\left(\frac{\Delta p}{p}\right)^2 \simeq \frac{\Delta p}{p}, \qquad \ldots\ldots(10.1)$$

and this is of the order of 10^{-4}. Also, using the fact that θ is small, we find

$$\begin{aligned} \hbar^2 k^2 &= p^2 + (p-\Delta p)^2 - 2p(p-\Delta p)(1-\theta^2/2) \\ &= p^2\left[\theta^2 + \left(\frac{\Delta p}{p}\right)^2 - \frac{\Delta p}{p}\theta^2\right]. \qquad \ldots\ldots(10.2) \end{aligned}$$

Thus, provided θ is greater than about 10^{-3} radian, yet still small enough for $\sin\theta$ to be replaced by θ, it is a very good approximation to write

$$p\theta = \hbar k. \qquad \ldots\ldots(10.3)$$

Using this expression in the dispersion relation (7.32), and neglecting the small term in k^4, we obtain the equation

$$\hbar\omega = \hbar\omega_p + \gamma\frac{p^2}{m}\theta^2, \qquad \ldots\ldots(10.4)$$

where

$$\gamma = 3\zeta_0/5\hbar\omega_p. \qquad \ldots\ldots(10.5)$$

Watanabe fitted equation (10.4) to his experimental data and obtained values of γ which agree very well with the theoretical values; in fact, the agreement is within the limits of experimental error for Be and Al, although rather worse for Mg and Ge. It is difficult to see how these results could be explained by the excitation of single electrons (cf. Ferrell 1956, Pines 1956).

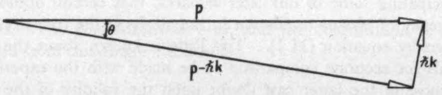


Figure 5. θ is the angle through which an incident electron of momentum $\mathbf{p}$ is scattered on emitting a plasmon of momentum $\hbar\mathbf{k}$.

We know that k has a maximum value, namely k_c, so we should expect the largest value of θ to be $\hbar k_c/p$. Watanabe, in fact, found the maximum scattering angle to be slightly greater than this theoretical value, if k_c is given by equation (7.30).

A discussion of the stopping power and the mean free path for plasmon emission would be out of place here. The subject has been reviewed in the article by Pines (1956), where further references may be found. Gabor (1956) has suggested new methods of distinguishing between plasma and single electron excitations, based upon the experimental techniques used in penetration experiments and their possible effects upon the observed energy spectra. In the following section it will be shown that further evidence of plasma oscillations in metals may be obtained from optical data.

§11. THE EFFECTS OF THE PERIODIC LATTICE FIELD AND OF THE CORE ELECTRONS

So far we have considered only the free-electron model of a metal, in which the ionic lattice is replaced by a uniform distribution of positive charge. It was pointed out in § 2 that this model is likely to give reasonable results for metals in which the ion cores are small, i.e. in which the core electrons are very tightly bound. However, even in these metals the periodic field of the ionic lattice must have *some* effect upon the frequency of the plasma oscillations, since it causes the one-electron wave functions to differ, in some degree, from free-electron functions. In cases where the cores are large, however, i.e. where the distinction between valence and core electrons tends to break down, we should expect a direct participation by the core electrons in the plasma oscillations, and hence a more serious divergence from the free-electron theory.

Taking the effect of the periodic field first, an obvious step which suggests itself is to replace the mass m in the expression for ω_p by the effective mass m^* of an electron in the periodic field. (N.B.—This has nothing to do with the effective mass introduced by the second canonical transformation, discussed in § 7.) It would take us too far out of our way to enter upon a discussion of the precise significance of the effective mass (but see equation (11.11) and the remarks which follow it). However, recognizing that m^* is in general a function of the wave vector $\mathbf{k}$ of a one-electron state, we might expect the frequency ω of plasma oscillation to be given by

$$\omega^2 = 4\pi ne^2\langle 1/m^*\rangle_{\text{av}} = \omega_p{}^2\langle m/m^*\rangle_{\text{av}}, \qquad \ldots\ldots(11.1)$$

the average being over all the occupied states. The alkali metals would provide the best test of this hypothesis, since for them m^* is approximately constant and easily calculated (cf. Bardeen 1938), but we have already said that the results of electron penetration experiments are unreliable for these metals; we might add, however, anticipating some of our later remarks, that certain optical data suggest that the frequency of plasma oscillation in the alkalis is the free-electron ω_p, rather than that given by equation (11.1). Too little is known about the effective mass in other metals for accurate comparison to be made with the experimental energy spectra, but most of the latter cast doubt upon the validity of the effective mass treatment. In the case of beryllium, for instance, the observed energy losses coincide *exactly* with those given by the free-electron plasma theory, although the work of Herring and Hill (1940) leads us to expect a value of $\langle m/m^*\rangle_{\text{av}}$ significantly different from unity. Furthermore, it is known that $\langle 1/m^*\rangle_{\text{av}}$ is zero for a full zone, so that the effective mass approximation leads to no plasma oscillation in insulators, which appears to contradict the results of electron penetration experiments on carbon, silicon and germanium.

The effective mass relation (11.1) was first derived by Wolff (1953) using a semi-classical treatment of the plasma oscillations, and the problem has been worked out in detail by Hubbard (1954) and Kanazawa (1955). These latter authors point out that the effective mass theory is essentially a *one-zone* theory and is only valid if the effect of virtual inter-zone transitions can be neglected, i.e. if the probability of an electron making a transition to another zone is negligibly small. Adams (1955 a) has given his opinion, however, that the effect of inter-zone transitions is likely to be so considerable that the effective mass treatment of plasma oscillations is not valid for any metal, an opinion which has the support of the anomalous experimental data mentioned above. We shall obtain a clearer view of the deficiencies of the effective mass treatment later on.

Knowledge of the band structure of metals is at present so scanty that a detailed theory of plasma oscillations, in which the effect of inter-zone transitions is correctly taken into account, is unlikely to yield accurate numerical results for some time to come. We shall therefore be content here to present a semi-quantitative theory due initially to Mott (1954). This has the advantage of showing simply and clearly how inter-zone transition probabilities affect the frequency of plasma oscillation, and leads to plausible explanations of many of the anomalies in the experimental energy loss spectra. Our method differs from that of Mott in some details, but the basic idea, of treating a plasma oscillation as a longitudinal polarization wave and applying what is essentially the Kramers–Heisenberg dispersion theory, remains unchanged.

Let us consider first the classical picture of a dielectric containing electrons which are bound to the atoms by Hooke's law forces. An electron which in the absence of an external field could execute simple harmonic motion with angular frequency ω_0 will, in the presence of a periodic electric field $\mathscr{E}_0 \cos \omega t$ along the x-axis, have the following equation of motion :

$$\frac{d^2x}{dt^2} + \omega_0{}^2 x = -\frac{e}{m}\mathscr{E}_0 \cos \omega t, \qquad \ldots\ldots(11.2)$$

where we are ignoring any damping. The solution of this equation consists of two parts : the first a simple harmonic vibration of frequency ω_0, which in practice would soon be damped out, and the second a particular solution

$$x = \frac{e/m}{\omega^2 - \omega_0{}^2}\mathscr{E}_0 \cos \omega t, \qquad \ldots\ldots(11.3)$$

due to the field. The *polarization* is $-ex$, and the contribution to the polarizability of the solid from this single electron is

$$\frac{-ex}{\mathscr{E}_0 \cos \omega t} = \frac{e^2/m}{\omega_0{}^2 - \omega^2}. \qquad \ldots\ldots(11.4)$$

Now, if f_i electrons per unit volume have a natural frequency ω_i, the total polarizability α of the solid is

$$\alpha = \frac{e^2}{m}\sum_i \frac{f_i}{\omega_i{}^2 - \omega^2}, \qquad \ldots\ldots(11.5)$$

the sum being over all the electrons in unit volume ; f_i is called the *oscillator strength* for frequency ω_i.

The quantum-mechanical dispersion theory of Kramers and Heisenberg (cf. Mott and Jones 1936, p. 103, Seitz 1940, p. 649) shows that a very similar formula gives the polarizability of a metal in a field of angular frequency ω, provided the oscillator strengths and the frequencies are suitably defined. For simplicity let us assume that the metal is optically isotropic, so that the polarizability is a scalar, and also that the occupied states are in a single Brillouin zone ; the *qualitative* results we shall obtain, however, should apply reasonably well to metals such as beryllium and magnesium, for which these assumptions are not true. The polarizability is then given by

$$\alpha = \frac{e^2}{m}\sum_{\mathbf{k}}\sum_{j}\frac{f_{\mathbf{k}j}}{\omega_{\mathbf{k}j}^2-\omega^2}, \qquad \ldots\ldots(11.6)$$

where **k** is the reduced wave vector of a one-electron state and j specifies a Brillouin zone; the **k**-sum is over all the occupied states (including spin) of the valence electrons per unit volume, and the j-sum is over all the zones. We define $\hbar\omega_{\mathbf{k}j}$ as the energy absorbed by an electron with wave vector **k** in making an optical transition from the occupied zone to zone j, and if we postulate that the occupied zone is that with $j = 0$, then $\omega_{\mathbf{k}0} = 0$ for all **k**. The oscillator strengths are defined as follows (Mott and Jones 1936):

$$f_{\mathbf{k}0} = 1-\frac{2}{\hbar m}\sum_{j\neq 0}\frac{|p_{\mathbf{k}}(j0)|^2}{\omega_{\mathbf{k}j}}, \qquad \ldots\ldots(11.7)$$

$$f_{\mathbf{k}j} = \frac{2}{\hbar m}\frac{|p_{\mathbf{k}}(j0)|^2}{\omega_{\mathbf{k}j}}, \quad \text{for} \quad j\neq 0, \qquad \ldots\ldots(11.8)$$

where

$$p_{\mathbf{k}}(j0) = \frac{\hbar}{i}\int\psi_{\mathbf{k}j}{}^*\frac{\partial}{\partial x}\psi_{\mathbf{k}0}\,d\tau, \qquad \ldots\ldots(11.9)$$

the field being in the x direction.

It follows from these definitions that

$$\sum_j f_{\mathbf{k}j} = 1. \qquad \ldots\ldots(11.10)$$

In order to avoid unnecessary complication we shall *define* the effective mass m^* by the equation

$$f_{\mathbf{k}0} = m/m^*, \qquad \ldots\ldots(11.11)$$

so that in general m^* depends upon **k**. This definition conforms with the more usual ones in those simple cases where the effective mass concept is most useful. For instance, when the one-electron energy $\epsilon(\mathbf{k})$ is given by

$$\epsilon(\mathbf{k}) = \epsilon(0)+bk^2, \qquad \ldots\ldots(11.12)$$

where b is a constant, which is generally assumed for the alkali metals, it can be shown (cf. Wilson 1953, p. 46) that

$$f_{\mathbf{k}0} = \frac{m}{\hbar^2}\frac{\partial^2\epsilon}{\partial k_x{}^2} = \frac{2mb}{\hbar^2}. \qquad \ldots\ldots(11.13)$$

Thus, using (11.11), we have $b = \hbar^2/2m^*$, and the last term of equation (11.12) has the same form as the kinetic energy for free electrons, equation (2.5), with m^* replacing m.

The result we require is now obtained by using the fact that $\alpha = -1/4\pi$ for a longitudinal wave. This may be proved by an extension of the treatment given in § 3. Recalling equation (3.1), we have

$$\operatorname{div}\mathscr{E} = 4\pi e(n-n'). \qquad \ldots\ldots(11.14)$$

The polarization current density $\mathbf{J}$ is given by

$$\mathbf{J} = \alpha \frac{\partial \mathscr{E}}{\partial t}, \qquad \text{......(11.15)}$$

and satisfies the equation of continuity,

$$-e\frac{\partial}{\partial t}(n-n') = \operatorname{div} \mathbf{J}. \qquad \text{......(11.16)}$$

Thus,

$$-e\frac{\partial}{\partial t}(n-n') = \alpha \operatorname{div} \frac{\partial \mathscr{E}}{\partial t} = \alpha 4\pi e \frac{\partial}{\partial t}(n-n'), \qquad \text{......(11.17)}$$

or

$$\alpha = -1/4\pi. \qquad \text{......(11.18)}$$

Substituting this result in (11.6), we obtain the following equation for the frequency ω of the plasma oscillations :

$$1 = \frac{4\pi e^2}{m} \sum_{\mathbf{k}} \sum_{j} \frac{f_{\mathbf{k}j}}{\omega^2 - \omega_{\mathbf{k}j}{}^2}, \qquad \text{......(11.19}a\text{)}$$

or

$$1 = \frac{4\pi e^2}{m}\left[\sum_{\mathbf{k}} \frac{f_{\mathbf{k}0}}{\omega^2} + \sum_{\mathbf{k}} \sum_{j \neq 0} \frac{f_{\mathbf{k}j}}{\omega^2 - \omega_{\mathbf{k}j}{}^2}\right], \qquad \text{......(11.19}b\text{)}$$

or

$$1 = \frac{4\pi e^2 n}{\omega^2}\left\langle \frac{1}{m^*} \right\rangle_{\mathrm{av}} + \frac{4\pi e^2}{m} \sum_{\mathbf{k}} \sum_{j \neq 0} \frac{f_{\mathbf{k}j}}{\omega^2 - \omega_{\mathbf{k}j}{}^2}. \qquad \text{......(11.19}c\text{)}$$

Equation (11.19*c*) shows us immediately what is neglected in the effective mass treatment. The equation cannot be solved directly, owing to our lack of knowledge about the $f_{\mathbf{k}j}$ and $\omega_{\mathbf{k}j}$, but certain deductions of a qualitative nature can be made.

(i) If $\omega \ll \omega_{\mathbf{k}j}$, for $j \neq 0$, the second term in (11.19*c*) may be negligible, giving the effective mass equation

$$\omega^2 = 4\pi e^2 n \langle 1/m^* \rangle_{\mathrm{av}}. \qquad \text{......(11.20)}$$

It is unlikely that this applies to any metal. As Adams (1955 a) points out, however, one might expect it to be true for plasma oscillations of the *conduction* electrons in semiconductors, i.e. the electrons in the conduction zone which are excess electrons coming from impurities or are due to temperature excitation of the electrons in the valence band. The density of these conduction electrons is small, so that ω will be small, and since only the lower part of the conduction zone is occupied it is reasonable to suppose that $\omega_{\mathbf{k}j}$ is large for all $j \neq 0$. It is tacitly assumed here that it is possible to treat the conduction and valence electrons separately, of course.

(ii) In the case of free electrons, we have $p_{\mathbf{k}}(j0) = 0$ and $f_{\mathbf{k}j} = 0$, for $j \neq 0$. Hence,

$$\sum_{\mathbf{k}} f_{\mathbf{k}0} = n, \qquad \text{......(11.21)}$$

and equation (11.19*b*) reduces to

$$1 = 4\pi e^2 n/m\omega^2, \qquad \text{......(11.22)}$$

i.e. $\omega = \omega_p$, as we found in § 3.

(iii) If, for all $\mathbf{k}$ and j, either $f_{\mathbf{k}j}$ is negligible or $\omega \gg \omega_{\mathbf{k}j}$, then from (11.19*a*) and (11.10) we obtain

$$1 \simeq \frac{4\pi e^2}{m\omega^2} \sum_{\mathbf{k}} \sum_{j} f_{\mathbf{k}j} = \frac{4\pi e^2 n}{m\omega^2}, \qquad \ldots\ldots(11.23)$$

i.e. $\omega \simeq \omega_p$. It should be noted that this does not necessarily mean that $m = m^*$, which may explain the free-electron plasma energy losses in beryllium, for instance. For this metal $\hbar\omega_p$ is 19 ev, which is much larger than the expected inter-zone transition energies for the zones immediately above the occupied ones, whose oscillator strengths should be appreciable. Hence, although $f_{\mathbf{k}0}$ differs from unity, the oscillation frequency may still be almost ω_p.

We have so far ignored the effect of the core electrons, except for their contribution to the field in which the valence electrons move. Their explicit effect on the frequency of plasma oscillation can be roughly represented by an additional term in equation (11.19*a*), which becomes

$$1 = \frac{4\pi e^2}{m}\left[\sum_{\mathbf{k}} \sum_{j} \frac{f_{\mathbf{k}j}}{\omega^2 - \omega_{\mathbf{k}j}^2} + \sum_{i} \frac{f_i}{\omega^2 - \omega_i^2}\right]. \qquad \ldots\ldots(11.24)$$

The extra term is taken straight from the classical theory of dielectrics, discussed above; we have merely assumed that f_i core electrons per unit volume have a natural frequency ω_i, and $\Sigma_i f_i$ is the total number of core electrons per unit volume.

If the core is small, i.e. if the core electrons are very tightly bound compared with the valence electrons, then $\omega_i \gg \omega$ and the effect of the core electrons on the plasma frequency of the valence electrons should be slight. This will be so, for instance, in the alkali metals and in beryllium, magnesium and aluminium.

In copper, however, the core is large, the s and d bands overlap, and so for some of the core electrons ω_i will no longer be very much larger than ω. In such cases no useful deduction can be made immediately from equation (11.24), although one might expect that there will not be a single frequency associated with the plasma oscillation, but a range of frequencies, so that the energy loss spectrum would show a broad band rather than a sharp peak. An alternative explanation of the broad loss band in copper has been proposed by Wolff (1953) and Herring (quoted by Wolff). They suggest that there is a large probability of a plasmon exciting an electron in the d band to a state above the Fermi level in the s band, so that the life of a plasmon is short and the line breadth consequently large.

In the case of an *insulator*, we might assume as a very crude model (Pines 1955) that all the electrons in the filled *valence* band, n per unit volume as before, have the same natural frequency ω_0. Then, if we ignore the tightly bound inner shells, and since there are no conduction electrons, equation (11.24) becomes

$$1 = \frac{4\pi e^2}{m} \frac{n}{\omega^2 - \omega_0^2}, \qquad \ldots\ldots(11.25)$$

or

$$\omega^2 = \omega_0^2 + \omega_p^2. \qquad \ldots\ldots(11.26)$$

It may well happen that $\omega_0 \ll \omega_p$, so that the frequency of oscillation, and consequent energy loss, would be that appropriate to a free-electron plasma with the valence electron density. This might explain the observed free-electron energy losses in carbon, silicon and germanium. We offer the alternative explanation, however, that case (iii) discussed above may apply well to these solids, for ω is large ($\hbar\omega_p$ is 25 ev for C, 17 ev for Si, and 16 ev for Ge) and quite possibly it is much larger than $\omega_{\mathbf{k}j}$ where $f_{\mathbf{k}j}$ is appreciable, particularly when the energy gap between valence and conduction bands is small, as it is for these intrinsic semiconductors.

Finally, we wish to mention briefly the connection between the plasma oscillations and the optical properties of metals, a subject which has been developed by a number of authors (Mott 1954, Hubbard 1955 a, b, Fröhlich and Pelzer 1955, Ferrell 1956, Pines 1956). Equation (11.18) tells us that the dielectric constant $1+4\pi\alpha$ vanishes at the frequency of plasma oscillation, so that we have here a means of determining the latter from optical data. Unfortunately, optical data in the relevant frequency range exist only for the alkali metals, but here the results are striking. The experiments of Wood (1933, cf. Zener 1933) show that for these metals the dielectric constant vanishes, i.e. the metals become transparent, for light of almost exactly the angular frequency ω_p. This strongly suggests that the free-electron plasma theory applies very well to the alkali metals, and that discrepancies in the energy loss spectra are due to experimental difficulties. Optical experiments on other metals, in the wavelength range roughly 500 Å to 1500 Å, would be very informative.

§ 12. CONCLUSION

We have seen that the effect of the long-range part of the Coulomb interaction of the electrons in a metal can be described in terms of plasma oscillations. Furthermore, the excitation energy of these plasma oscillations is so great that they are normally in their ground states and take no active part in electronic processes; in many calculations, therefore, the long-range part of the Coulomb interaction may be ignored. The remaining interaction has a range of only about one ångström unit, so that as a first approximation the electrons may be regarded as non-interacting. This explains the many successes of the elementary free-electron and band theories of metals; in addition to the examples cited in the text, it explains, for instance, how it is possible to ignore electron–electron scattering in conduction theory (Abrahams 1954, Bardeen and Pines 1955). We have seen that this re-description of the electronic motion in terms of plasma oscillations reduces the exaggerated 'exchange' effects associated with the Hartree–Fock theory, and yet, when the zero-point energies of the oscillations are taken into account, it gives very good results in cohesive energy calculations. Finally we have seen that the plasma theory provides a rational explanation of some of the energy losses observed in electron penetration experiments, and is in accord with such relevant optical data as are available. It is true that these experimental results are not conclusive; some explanation of them independent of plasma oscillations may well be found, but none as successful as the plasma theory has yet been forthcoming. Detailed calculation for actual metals, going beyond the free-electron approximation, is badly needed here.

The results of the plasma theory have now been used with some success in many calculations which we have been unable to mention, references to which may be found in the several review articles we have quoted. Our main purpose has been to make the basis of the theory plausible, by elementary arguments wherever possible. The quantum-mechanical theory we have described is not free from difficulties, but it is not too much to hope that minor adjustments may remedy these. In any case, whether or not large-scale revision will prove necessary, it cannot be denied that the theory of plasma oscillations is a stimulating one and offers a new approach to the difficult problems raised by the interaction of electrons in metals.

ACKNOWLEDGMENTS

I am indebted to Professor H. Jones for several helpful suggestions during the writing of this article, as well as for reading and commenting upon the completed manuscript. I have had rewarding discussions of particular topics with Dr. D. Pines, Dr. J. Hubbard and Dr. D. Gabor, and I am grateful to Dr. A. Seeger for drawing my attention to the paper by Steenbeck (1932), which is of historical interest. Finally, I wish to thank all those authors who have sent me reprints of their papers, some of which I have been unable to mention owing to lack of space.

REFERENCES

ABRAHAMS, E., 1954, *Phys. Rev.*, **95**, 839.

ADAMS, E. N., 1955 a, *Phys. Rev.*, **98**, 947 ; 1955 b, *Ibid.*, **98**, 1130.

BARDEEN, J., 1936, *Phys. Rev.*, **50**, 1098 ; 1938, *J. Chem. Phys.*, **6**, 367.

BARDEEN, J., and PINES, D., 1955, *Phys. Rev.*, **99**, 1140.

BLACKSTOCK, A. W., RITCHIE, R. H., and BIRKHOFF, R. D., 1955, *Phys. Rev.*, **100**, 1078.

BOHM, D., and PINES, D., 1951, *Phys. Rev.*, **82**, 625 ; 1952, *Ibid.*, **85**, 338 ; 1953, *Ibid.*, **92**, 609.

BOHR, A., 1948, *K. Danske Vidensk. Selskab, Mat.-fys. Medd.*, **18**, Nr. 19.

BROOKS, H., 1953, *Phys. Rev.*, **91**, 1027.

DEBYE, P., and HÜCKEL, E., 1923, *Phys. Z.*, **24**, 185.

FERRELL, R. A., 1956, *Phys. Rev.*, **101**, 554.

FOCK, V., 1930, *Z. Phys.*, **61**, 126.

FRÖHLICH, H., and PELZER, H., 1955, *Proc. Phys. Soc.* A, **68**, 525.

GABOR, D., 1956, *Phil. Mag.* [8], **1**, 1.

HARTREE, D. R., 1928, *Proc. Camb. Phil. Soc.*, **24**, 89.

HERRING, C. C., and HILL, A. G., 1940, *Phys. Rev.*, **58**, 132.

HUBBARD, J., 1954, *Proc. Phys. Soc.* A, **67**, 1058 ; 1955 a, *Ibid.*, **68**, 441 ; 1955 b, *Ibid.*, **68**, 976.

JONES, H., 1954, *Phys. Rev.*, **94**, 1072.

KANAZAWA, H., 1955, *Progr. Theor. Phys., Japan*, **13**, 227.

KRONIG, R., and KORRINGA, J., 1943, *Physica*, **10**, 406, 800.

KUPER, C. G., 1956, *Proc. Phys. Soc.* A, **69**, 492.

LANG, W., 1948, *Optik*, **3**, 233.

LANGMUIR, I., 1928, *Proc. Nat. Acad. Sci., Wash.*, **14**, 627.

MACKE, W., 1950, *Z. Naturf.*, **5a**, 192.

MARTON, L., LEDER, L. B., and MENDLOWITZ, H., 1955, *Advances in Electronics and Electron Physics* (New York : Academic Press), p. 183.

MOTT, N. F., 1954, *Proceedings of the Tenth Solvay Congress*, Brussels.

MOTT, N. F., and JONES, H., 1936, *Theory of the Properties of Metals and Alloys* (Oxford : University Press).

PINES, D., 1953, *Phys. Rev.*, **92**, 626 ; 1954, *Proceedings of the Tenth Solvay Congress*, Brussels ; 1955, *Solid State Physics*, **1** (New York : Academic Press), p. 367 ; 1956, *Rev. Mod. Phys.*, **28**, 184.

RAIMES, S., 1950, *Phil. Mag.* [7], **41**, 568 ; 1952, *Ibid.*, **43**, 327 ; 1953, *Proc. Phys. Soc.* A, **66**, 949 ; 1954 a, *Ibid.*, **67**, 52 ; 1954 b, *Phil. Mag.* [7], **45**, 727 ; 1954 c, *Proceedings of Symposium on Band Structure*, Radar Research Establishment, Malvern, p. 16 ; 1956, *Research*, **9**, 374.

Ruthemann, G., 1941, *Naturwissenschaften*, **29**, 648; 1942, *Ibid.*, **30**, 145; 1948, *Ann. Phys., Lpz.*, **2**, 113.
Seitz, F., 1940, *Modern Theory of Solids* (New York: McGraw-Hill).
Skinner, H. W. B., 1940, *Phil. Trans. Roy. Soc.* A, **239**, 95.
Slater, J. C., 1939, *Introduction to Chemical Physics* (New York: McGraw-Hill); 1953, *Rev. Mod. Phys.*, **25**, 199.
Sommerfeld, A., 1928, *Z. Phys.*, **47**, 1.
Steenbeck, M., 1932, *Z. Phys.*, **76**, 260.
Tonks, L., and Langmuir, I., 1929, *Phys. Rev.*, **33**, 195.
Watanabe, H., 1956, *J. Phys. Soc., Japan*, **11**, 112.
Wigner, E., 1934, *Phys. Rev.*, **46**, 1002; 1938, *Trans. Faraday Soc.*, **34**, 678.
Wigner, E., and Seitz, F., 1933, *Phys. Rev.*, **43**, 804; 1934, *Ibid.*, **46**, 509.
Wilson, A. H., 1953, *Theory of Metals*, 2nd edn. (Cambridge: University Press).
Wohlfarth, E. P., 1950, *Phil. Mag.* [7], **41**, 534.
Wolff, P. A., 1953, *Phys. Rev.*, **92**, 18.
Wood, R. W., 1933, *Phys. Rev.*, **44**, 353.
Zener, C., 1933, *Nature, Lond.*, **132**, 968.

THE THEORY OF LIQUID ^{4}He

By J. WILKS

Clarendon Laboratory, Oxford

CONTENTS

PAGE
§ 1. Introduction 39
§ 2. Phonons 41
§ 3. Rotons 42
§ 4. The two-fluid model 47
§ 5. Applications of the two-fluid model 50
  5.1. Superfluidity 50
  5.2. The normal viscosity and Andronikashvili's experiment 51
  5.3. The fountain and mechano-caloric effects 53
  5.4. Heat transport 55
  5.5. Temperature waves 56
  5.6. The momentum of heat flow 58
§ 6. The excitation gas 59
  6.1. The viscosity of the normal fluid 60
  6.2. The attenuation of pressure and temperature waves 61
  6.3. Mean free path effects 62
§ 7. Turbulence 64
  7.1. Viscosity and heat flow experiments 65
  7.2. The critical velocity 67
  7.3. Vortex lines 69
  7.4. Rotating helium 70
§ 8. The helium film 70
  8.1. The thickness of the film 71
  8.2. Transport properties 72
  8.3. The unsaturated film 73
§ 9. The permanent liquids 74
  9.1. Liquid helium I 74
  9.2. Liquid ^{3}He 75
  9.3. Liquid ^{3}He–^{4}He mixtures 76
  9.4. Statistics and the lambda point 79
§ 10. Conclusion 80
  References 80
  Appendix 83

Abstract. An account is given of the present state of the theory of liquid ^{4}He. At low temperatures it is no longer possible to consider the motion of individual atoms, and a wave mechanical treatment becomes necessary. Such a treatment, as developed by Landau and Feynman, characterizes the thermal motion of the liquid by 'elementary excitations', the so-called phonons and rotons. A consideration of these excitations then leads to the 'two-fluid' model. It is shown that the treatment is adequate to account for nearly all the properties of the liquid, which are briefly reviewed. The principal problem still outstanding concerns the behaviour of the liquid at high rates of flow or when carrying a large heat current, when it is clear that some form of turbulence is induced.

§ 1. INTRODUCTION

IN this review we shall confine our attention principally to liquid helium II, that is to liquid ^{4}He at temperatures below the so-called lambda point at 2·17°K, which is characterized by an anomaly in the specific heat (figure 1). Above this temperature the liquid is not vastly different from many others; below this temperature it is unique and exhibits quantum effects on a macroscopic scale. Considerable contributions to the theory have been made by F. London, Tisza, Landau and Feynman; in particular, Landau in his 1941 paper gave most of the picture as we see it to-day. However, some of his arguments were not presented in any great detail, and as one section of the paper dealing with quantum hydrodynamics is now recognized to be incorrect, the unusual concepts which he introduced did not find immediate acceptance. However, an increasing number of experimental results and theoretical speculations have justified his approach; also the results he deduced by arguments involving quantum hydrodynamics have been more satisfactorily established by Feynman. This review attempts to set down the present position in a logical manner, both in order to give a general picture to the reader not intimately acquainted with this field, and also to delineate those points where the theory still appears inadequate or in need of modification. In the first four sections we describe the quantum concepts necessary for the treatment of liquid helium, and then in the later sections apply these results to the experimental observations. In order to present as clear a picture as possible, no attempt is made to refer to all the published material, but this is hardly necessary as several comprehensive reviews covering most aspects of the subject are already available (e.g. Dingle 1952 a, Daunt 1952, Wilks 1953, Daunt and Smith 1954). However, in the appendix we give a list of the papers on liquid ^{4}He which have appeared since these reviews were published, and which are not referred to in the present article.

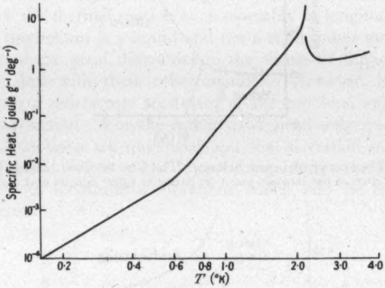


Figure 1. The specific heat of liquid helium (*after* Kramers, Wasscher and Gorter 1952, and Keesom and Clusius 1932).

As may be seen from the melting curve shown in figure 2 (Simon and Swenson 1950), the most important single aspect of liquid helium is that, under its own vapour pressure, it remains a liquid down to absolute zero, and will only solidify under a pressure of at least 25 atmospheres. This behaviour is a consequence of the van der Waals forces being very small; on considering the equilibrium spacing of the atoms in the condensed phase, it is found that the energy associated with the zero-point motion is comparable with the ordinary potential and kinetic-energy terms (Simon 1934). The effect of this zero-point energy is to increase

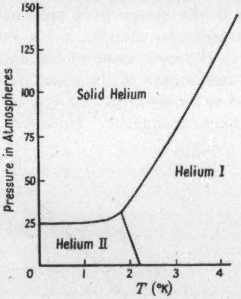


Figure 2. The melting curve of liquid helium. The line between helium I and helium II gives the dependence of the lambda point on pressure (*after* Simon and Swenson 1950).

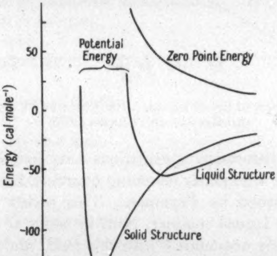


Figure 3. The internal energy of condensed helium at 0°K. The lowest total energy, consisting of the sum of potential and zero-point energy, is obtained with the liquid phase at a molar volume of 27 cm^3 (*after* F. London 1954).

the net kinetic energy, and therefore to increase the atomic spacing to such an extent that the atomic volume is about three times as large (27 cm^3) as would be expected in the absence of zero-point energy. Hence the atoms never come close enough for the van der Waals forces to overlap and form a solid phase. It is difficult to make an exact calculation of the equilibrium condition of a system in which zero-point energy plays a decisive role, but a full discussion of what is possible has been given by F. London (1954). The potential energy of both liquid and solid helium may be calculated from the virial coefficients together with a knowledge of the form of the respective lattices obtained from X-ray analyses (Keesom and Taconis 1938 a, b); the results for temperature zero are shown in figure 3, together with an estimate of the zero-point energy. It can be seen that the total energy at absolute zero, which consists only of potential plus zero-point energy, is least for liquid with a molar volume of about 27 cm³; the liquid at this volume has less energy than the solid and is therefore the stable phase.

§ 2. PHONONS

Given that the condensed phase of helium at absolute zero is a liquid, we consider what thermal motions are introduced as the temperature is raised. It is an experimental fact that the liquid will transmit longitudinal sound waves (although not transverse ones on account of its very low viscosity), and one therefore concludes that part of the thermal motion is an assembly of longitudinal sound waves. Obviously, since the helium is a liquid and not a solid, other modes of motion are possible, and one of the great difficulties in the theory of liquid helium has been to know how to deal with these other motions. However, it now seems that Feynman has given a satisfactory treatment of the problem, and this will be discussed in the next section. For the moment we need only quote the result that the other thermal motions are quantized and that a certain minimum energy is needed to excite them. Thus at the lowest temperatures the heat motion of the liquid is due only to longitudinal vibrations which may be expected to have a Debye-type spectrum.

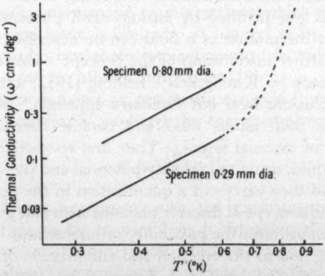


Figure 4. The thermal conductivity of liquid helium at low temperatures. Below 0·6°K the heat flow is proportional to the temperature gradient, and the apparent thermal conductivity varies as the diameter of the specimen (*after* H. A. Fairbank and J. Wilks 1955).

The conclusion that no other modes of motion are permitted, although rather surprising, is well supported by the experimental evidence. Thus below about 0·6°K the specific heat (figure 1) may be written as $0{\cdot}0235T^3$ joules g^{-1} deg^{-1} (Kramers, Wasscher and Gorter 1952), and this is in quite good agreement with the value of $0{\cdot}021T^3$ joules g^{-1} deg^{-1} calculated from the Debye theory using a value of the velocity of sound obtained by extrapolating the results of Atkins and Chase (1951). The ten per cent discrepancy in the numerical factor is of no great significance; it may be due to some errors in the rather difficult experiments, or perhaps could be explained by a more refined treatment of the sound waves as has been attempted for instance by Lokken (1954). Figure 4 gives values of the thermal conductivity (Fairbank and Wilks 1955) which show that although the behaviour of the liquid is very complex above 0·6°K, below that temperature the liquid behaves in a very similar way to a dielectric solid. The heat flow per unit area is accurately proportional to the temperature gradient, and is given quite closely by the formula of Casimir (1938) for the thermal conductivity of a dielectric solid in which the heat flow is limited only by scattering at the boundary of the specimen. Thus these measurements demonstrate both the presence of the thermal waves and also the absence of any other form of motion capable of scattering these waves in such a way as to produce thermal resistance. Finally the coefficient of thermal expansion, which becomes negative immediately below the lambda point, appears at low temperatures to take up positive values in good agreement with what would be expected if only sound waves were present (Atkins and Edwards 1955).

There is one further point essential to the treatment of the heat motion associated with the longitudinal sound waves: as we are concerned with very low temperatures a quantum treatment is necessary and was first introduced by Landau (1941). He remarks that any weakly excited state can be considered as an aggregate of single elementary excitations which in this particular case are sound quanta, or phonons, with discrete values of energy ϵ and of momentum p related by

$$\epsilon = pc.\dagger \qquad \ldots\ldots(1)$$

The concept of phonons to describe a sound field seems to have been first mentioned by Frenkel (1936) and was justified by analogy with photons. Although it is by no means obvious that the motion of a fluid can be described in this way, Landau did not attempt any further justification of the concept. However, a full investigation has since been made by Kronig and Thellung (1952) and also by F. London (1954), and their conclusions bear out Landau's approach. Kronig and Thellung consider a continuous non-viscous fluid and confine themselves to irrotational motion in the absence of external forces. They first re-state the classical equations of motion for such a liquid, using the velocity potential and the density as canonically conjugate variables, and then carry out a quantization in the normal manner. This leads immediately to the concept of discrete phonons with energy ϵ and momentum p. The treatment also demonstrates the possibility of interactions between the phonons giving rise to scattering and to the creation and annihilation of phonons—processes which had already been postulated and discussed by Landau and Khalatnikov (1949).

§ 3. ROTONS

As is shown in figure 1, the specific heat above 0·6°K becomes rapidly greater than the Debye value, and it is clear that some other forms of thermal motion are being excited. The question, however, is just what form these excitations take. Landau believed that some form of vortex or rotational motion was involved, and to investigate this possibility set up a new formulation of quantum hydrodynamics. He claimed to show that vortex motion is quantized, and that the energy and

---

† where c is the velocity of sound.

momentum of an elementary excitation or 'roton' are related by the equation

$$\epsilon = \Delta + p^2/2\mu \qquad \ldots\ldots(2)$$

where Δ and μ are constants. This relation shows that a minimum energy Δ is required to create a roton, and immediately leads to the conclusion that at sufficiently low temperature the rotons will not contribute to the thermal energy. At higher temperatures the total thermal energy of the liquid is the energy of an assembly of photons and rotons, and the thermodynamic functions may be calculated by the usual statistical methods. By comparing the expressions for the specific heat obtained in this way with the experimental values, Landau obtained values for Δ/k of 8–9°K (k is Boltzmann's constant), and for μ of 7–8 times the mass of the helium atom. This comparatively large value for Δ correlates with the absence of rotons below about 0·6°K and the rapid rise in the specific heat above that temperature.

Although the concept of a roton with an energy spectrum similar to that of equation (2) has turned out to be very fruitful, Landau's derivation of it is now regarded as unacceptable. It is far from clear that the operators in his quantum-mechanical treatment have the properties he somewhat intuitively claims for them (for detailed criticisms, see F. London 1945 and Dingle 1952 a). Quite recently Ziman (1953 a) made a new attack on the problem rather in the way that Kronig and Thellung (1952) had treated the sound waves; he was able to set up the classical equations of hydrodynamics in a form suitable for quantization and claimed to show that the Hamiltonian contained terms corresponding to the excitation of a roton state. However, this treatment also appears open to objection, because other terms in the Hamiltonian correspond to interactions which have to be taken into account if more than one roton is excited. It has yet to be shown that the same conclusions will be valid when these interactions are included; other difficulties have also been pointed out by Allcock and Kuper (1955).

After Peshkov's measurements on temperature waves (to which we refer later), Landau (1947) quite empirically modified the energy spectrum to the form:

$$\epsilon = \Delta + \frac{(p-p_0)^2}{\mu} \qquad \ldots\ldots(3)$$

in order to obtain better agreement with the experimental results. This last equation has turned out to be so useful in interpreting the behaviour of liquid helium that the uncertain foundations on which it was based were rather glossed over until recently, when Feynman gave a derivation of the equation which is far more satisfactory. Feynman also shows, what Landau only tacitly assumed, that the phonon states are the only possible states at low temperatures. (As Landau's hydrodynamical approach used only coarse-grained variables, it did not rule out the possibility of other excitations associated with motion on an atomic scale.) Finally, before giving a summary of Feynman's treatment, it should be pointed out that although Landau's derivation of the roton spectrum is almost certainly incorrect, it was he who made the essential step of postulating that the whole motion of the liquid could be expressed in terms of quantized elementary states. These states are, it must be emphasized, wave-mechanical states of the whole liquid and are characterized by their energy and momentum.

We now give a brief summary of Feynman's approach, referring the reader to the original papers for a full and lucid discussion of the ideas involved (Feynman 1953 a, b, 1954, 1955; Feynman and Cohen 1956). He regards the liquid as an assembly of atoms, between pairs of which there are strong repulsive and weak attractive van der Waals forces, and which to a first approximation may be treated as hard spheres of diameter about $2{\cdot}7A$. As the mean volume per atom is about $(3{\cdot}6A)^3$, the structure is a very open one, and the atoms can move about freely either through holes in the lattice or by pushing other atoms aside. Feynman first discusses density fluctuations, and agrees with previous authors that there will be a continuous range of energy levels right down to the lowest values, corresponding to sound waves of different wavelengths. He next considers what other forms of excitation are possible, and remarks that any such excitations will correspond to various rearrangements of the atoms at constant density. Moreover, because of the very open structure of the liquid, such rearrangements may in principle be brought about by a simple stirring process, and Feynman proceeds to deduce appropriate wave functions to describe the liquid. The wave functions used in his first papers were recognized to be only approximate and led to difficulties concerning the conservation of mass (Feynman 1954), and better functions have been given recently by Feynman and Cohen (1956). However, as the earlier work forms a better introduction to the subject, and as no great difference of principle is involved, we now outline Feynman's theory using his earlier wave functions.

Suppose the helium to be in the rth stationary state with a given energy E_r and let the position of the ith atom be defined by its coordinate $\mathbf{R}_i$; then the wave function for that state may be written $\psi_r(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N)$, a definite though complicated function of the coordinates $\mathbf{R}$. As the atoms have a tendency to keep apart, ψ_r will have its greatest values for those configurations in which the atoms are widely separated; and as the atoms are regarded as impenetrable spheres ψ_r must fall to zero for configurations in which atoms overlap. As usual the ground state $\psi = \phi$ is characterized by the absence of nodes, and ϕ may be taken to be a positive number for all configurations of the $\mathbf{R}$'s.

Now consider a wave function ψ_r which is to represent a stationary state of higher energy. ψ_r must be orthogonal to the ground state ϕ, which is everywhere positive; that is, ψ_r must be positive for some configurations $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_N)$ and negative for the others. Thus if ψ_r is plotted against the $3N$ coordinates of configuration space it will vary between certain maximum positive and negative values. We wish to find states of low energy, and the criterion for such states is that the gradient of ψ_r in configuration space should be small. (We recall that in the simple problem of a particle in a box, the wave function oscillates more frequently the higher the energy, as the kinetic energy is given by $-(\hbar^2/2m)\nabla^2\psi$.) Let us now refer to the configuration which gives ψ_r its maximum positive value as an A configuration, and that which gives the maximum negative value as a B configuration. Then if ψ_r is to represent a state of low energy, the points corresponding to configurations A and B should be widely separated in configuration space. If the atoms are distinguishable, this condition implies that to move the atoms from an A configuration to a B configuration involves a stirring process in which all the atoms are displaced a long way. However, as all the atoms are indistinguishable and obey Bose statistics, it must be possible to go from an A configuration to a B, or indeed to any other one, by moving each atom by a distance of no more than about the atomic spacing. This means that configurations A and B can only be separated by a certain limited distance in configuration space, the energy cannot be made arbitrarily small and therefore will not fall below a certain minimum value. Thus at sufficiently low temperatures the probability of exciting such states is very low and the only excitations will be phonons.

Feynman now deduces an explicit form for the wave function ψ_r and hence determines the energy spectrum of the other excitations. As defined above, ψ_r has a maximum value for configuration A and a minimum value for configuration B, and for low energy states it must vary smoothly in going from A to B. To ensure this, the amplitude of ψ_r is taken to be the number of atoms on sites of A-type configurations minus the number on B configuration sites. This number is $\Sigma f(\mathbf{R})$ where $f(\mathbf{R})$ is a function which is $+1$ if $\mathbf{R}$ defines an A configuration site and -1 if $\mathbf{R}$ defines a B position, and which varies smoothly between these limits. Combining this with the condition that atoms must not overlap, the wave function is written

$$\psi_r = \sum_i f(\mathbf{R}_i)\phi$$

where ϕ, the wave function of the ground state, is introduced to ensure that the atoms do not overlap and to take account of correlations between the atoms. Although Feynman does not specify the motion of the atoms more precisely than by this last equation, he points out that his wave function is consistent with three quite probable modes of motion, viz. the rotation of a small ring of atoms, the excitation of single atoms in a cage, and single atoms vibrating with wave number about $2\pi/a$ where a is the atomic spacing, the surrounding atoms moving about to get out of their path. Feynman then applies the variational principle to find the form of $f(\mathbf{R})$ which gives the lowest energy and thus obtains

$$f(\mathbf{R}) = \exp(i\mathbf{k}.\mathbf{R})$$

and

$$\epsilon(k) = \hbar^2 k^2/2m\,S(k) \qquad \ldots\ldots(4)$$

where S is the Fourier transform of the correlation function

$$S(\mathbf{k}) = \int p(\mathbf{R})\exp(i\mathbf{k}.\mathbf{R})\,d^3\mathbf{R}.$$

The correlation function, or liquid structure factor, may be deduced either from x-ray data (Reekie and Hutchison 1953) or from neutron diffraction data (Henshaw and Hurst 1953), and the form of the function is shown in figure 5 (*a*). Figure 5 (*b*) shows the resulting curve for the energy given by equation (4), the upper part being similar to Landau's roton spectrum (equation (3)). The height of the minimum of the curve is somewhat high, corresponding to too large a value of Δ; however, the more exact wave functions of Feynman and Cohen (1956) lead to a value for Δ/k of 11·5°K, in much better agreement with the value of about 9·6°K derived from experimental results. We may mention here that the values of Δ, μ and p_0 are best determined from a consideration of the magnitude and temperature dependence of the specific heat and of the 'normal density', to which we refer later. Due to experimental uncertainties, plus of course the fact that the roton spectrum itself is almost certainly only an approximation, it is difficult to obtain a quite definite answer; however, reasonable agreement results if one takes $\Delta/k = 9{\cdot}6^\circ\text{K}$, $p_0/\hbar = 2{\cdot}0 \times 10^{-8}$ and $\mu = 0{\cdot}77m$, where m is the mass of the helium atom (de Klerk, Hudson and Pellam 1954).

The low energy part of the spectrum corresponds to the relation $\epsilon = pc$ and is just the phonon spectrum discussed previously. At first sight this is somewhat surprising as figure 5 (*b*) then suggests that the energy of an excitation can have any value between zero and Δ. This point is still not entirely clear, but Feynman (1955) remarks that the wave functions deduced by the variational method are in general only appropriate in the region of minimum energy. It happens that his wave function for the rotons also turns out to be an exact wave function for phonons of small wave number, and the method therefore gives the lower end of the spectrum correctly. However, this does not imply that the wave function is good for intermediate values of the energy, and therefore phonons and rotons may be quite distinct modes of motion. Indeed, Feynman and Cohen claim that their latest wave function for rotons corresponds roughly to a vortex ring of such small radius that only one atom can pass through its centre, outside the ring there being a slow drift of atoms returning for another passage through the ring.

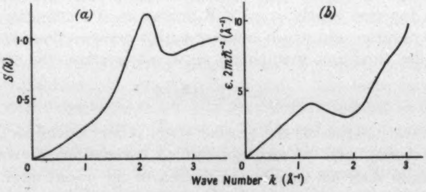


Figure 5. (*a*) The structure factor for liquid helium as determined by neutron diffraction (*after* Henshaw and Hurst 1953); (*b*) The energy spectrum for excitations in liquid helium, as calculated by Feynman 1954.

It is to be noted that the above theory does not account, as it presumably should, for the fact that, between about 1°K and the lambda point, the liquid exhibits a negative expansion coefficient. It is possible to relate this coefficient to the three parameters Δ, p_0 and μ, and their dependence on the density of the helium (see, for example, Atkins 1955), but Feynman has not as yet discussed what dependence on density is to be expected. On the other hand, Franchetti (1954) has given an explanation of the negative expansion coefficient in terms of exchange forces; and Prigogine and his collaborators have given an explanation based on a cell model in which some cells may be doubly occupied and in which the part played by zero-point energy is predominant (Prigogine 1954). The variation of the parameters Δ, p_0 and μ with the density of the helium is also of interest in that it offers, at least in principle, a further test of the theory. Under a pressure of say 25 atmospheres the density is changed by about 10% and we have a liquid to which the theory should still apply but with appreciably different parameters; thus the lambda point is depressed to about 1·8°K, the specific heat is increased by about 50% (Hercus and Wilks 1955), and the velocities of sound and temperature waves are changed considerably (Atkins and Stasior 1953, Maurer and Herlin 1951).

However, at present some of the data essential for a complete comparison is not available; for example there are as yet no direct determinations under pressure of the 'normal' density (see § 4), nor of the correlation function.

§ 4. THE TWO-FLUID MODEL

We now give the derivation of the remarkable two-fluid model; this derivation was first given by Landau but with several of the intermediate steps omitted, and the position has since been clarified by Dingle, F. London, and particularly by Feynman. Before giving a summary of these arguments, we should perhaps mention that they apply equally well to any form of excitations with given energy and momenta, and do not depend on particular properties of phonons or rotons.

We begin by calculating the total momentum of a mass of helium containing excitations and moving with a constant velocity. To carry out this calculation we need the result that if the energy of an excitation in the liquid at rest is ϵ then the energy that an observer at rest must associate with the same excitation in liquid moving with velocity $\mathbf{v}$ is given by

$$\epsilon_0 = \epsilon + \mathbf{p}.\mathbf{v}. \qquad \ldots\ldots(5)$$

This important result was first deduced by Landau from a consideration of a Galilean transformation in classical mechanics; we give here a rather similar but clearer derivation due to Dingle (1952 a). For an observer who is moving with the liquid and whose coordinate system is indicated by primes, the wave function of an excitation in the liquid may be written

$$\psi = A \exp\left\{\frac{1}{i\hbar}(\epsilon_0 t' - \mathbf{p}.\mathbf{r}')\right\}.$$

The appropriate wave function for an observer at rest (unprimed coordinates) is obtained by the usual transformation

$$\mathbf{r} = \mathbf{r}' + \mathbf{v}t. \quad (t = t')$$

Hence for the observer at rest

$$\psi = A \exp\left\{\frac{1}{i\hbar}[\epsilon t - \mathbf{p}.(\mathbf{r} - \mathbf{v}t)]\right\}$$

$$= A \exp\left\{\frac{1}{i\hbar}[(\epsilon + \mathbf{p}.\mathbf{v})t - \mathbf{p}.\mathbf{r}]\right\}.$$

Thus we have the required result that if the fluid is moving with velocity $\mathbf{v}$ relative to an observer, the energy of an excitation appropriate to his system of coordinates is $\epsilon + \mathbf{p}.\mathbf{v}$. It is perhaps worth mentioning, in view of the arguments shortly to be used, that this result is based purely on mechanical considerations, as yet statistics have played no part in the discussions.

We now consider moving liquid containing a large number of excitations, and apply statistics to determine the total momentum of the assembly. It is convenient to specify that the excitations have an arbitrary drift velocity $\mathbf{v}_n$ with respect to the observer. To fix our ideas consider a cylinder of liquid moving with velocity $\mathbf{v}_s$ and filling a tube moving with velocity $\mathbf{v}_n$, the excitations being in equilibrium with the walls of the tube. Statistical mechanics gives the equilibrium distribution in a gas which has a uniform drift velocity $\mathbf{v}_n$ relative to an observer as

$$n(\epsilon) = [\exp \beta(\epsilon_0 - \mathbf{p}\,.\,\mathbf{v}_n) - 1]^{-1}.$$

Treating the excitations as gas-like particles and substituting for ϵ_0 from equation (5) we obtain

$$n(\epsilon) = [\exp \beta\{(\epsilon + \mathbf{p}\,.\,\mathbf{v}_s) - \mathbf{p}\,.\,\mathbf{v}_n\} - 1]^{-1}.$$

The net momentum associated with the excitations is not zero, as it would be if the liquid were at rest, and is given by

$$\bar{\mathbf{p}} = \int \mathbf{p}\{\exp \beta[\epsilon(\mathbf{p}) + \mathbf{p}\,.\,\mathbf{v}_s - \mathbf{p}\,.\,\mathbf{v}_n] - 1\}^{-1}(2\pi\hbar)^{-3}\,d\mathbf{p}^3.$$

Making the valid assumption that $\mathbf{v}_s$ and $\mathbf{v}_n$ are small, and expanding to the first order in $\mathbf{v}_s - \mathbf{v}_n$ we find

$$\bar{\mathbf{p}} = -\rho_n(\mathbf{v}_s - \mathbf{v}_n)$$

where ρ_n, an essentially positive quantity with the dimensions of density, is defined by

$$\rho_n = -\tfrac{1}{3}\beta \int p^2[\exp \beta\epsilon(p) - 1]^{-2} \exp \beta\epsilon(p)\,(2\pi\hbar)^{-3}\,d^3\mathbf{p}. \qquad \ldots\ldots(6)$$

Note that in this linear approximation ρ_n will be a function of temperature but not of $\mathbf{v}_n$ or $\mathbf{v}_s$. The total momentum $\mathbf{j}$ associated with unit volume of moving liquid will be the ordinary $m\mathbf{v}$ term plus the momentum associated with the excitations. Thus

$$\mathbf{j} = \rho\mathbf{v}_s + \bar{\mathbf{p}} = \rho\mathbf{v}_s - \rho_n(\mathbf{v}_s - \mathbf{v}_n).$$

Defining a new parameter ρ_s by

$$\rho_s + \rho_n = \rho \qquad \ldots\ldots(7)$$

we have

$$\mathbf{j} = \rho_s\mathbf{v}_s + \rho_n\mathbf{v}_n. \qquad \ldots\ldots(8)$$

The last equation implies that the total momentum of the moving fluid may be formally resolved into two components, one associated with the excitations and the other with the remainder of the fluid. Clearly from the statistical manner in which the quantity ρ_n is derived it is not to be associated with the mass of particular atoms in the system. (This is underlined by the fact that below 0·6°K the values of the specific heat and its T^3 dependence on the temperature indicate that *all* the atoms have vibrational energy—at least in the classical sense—whereas ρ_n turns out to have a value which is only 10^{-4} or less of the density of the liquid.) Nevertheless the terms $\rho_s\mathbf{v}_s$ and $\rho_n\mathbf{v}_n$ are the momenta associated with the motion of the 'background' and of the excitations respectively and thus ρ_s and ρ_n play a role analogous to that of ordinary inertial mass. This analogy finds further support when one calculates the internal energy of a fluid containing excitations; it is found that the internal energy of fluid for which $\mathbf{v}_s$ and $\mathbf{v}_n$ have finite values is greater than that for fluid with zero values of $\mathbf{v}_s$ and $\mathbf{v}_n$, by an amount $\frac{1}{2}\rho_s v_s^2 + \frac{1}{2}\rho_n v_n^2$. Thus ρ_s and ρ_n again behave as inertial masses and this behaviour is the basis of the so-called two-fluid model of helium, in which the motion of the liquid is formally divided into two parts: that of the 'normal fluid', which has 'density' ρ_n and velocity $\mathbf{v}_n$, and that of the 'superfluid', which has 'density' ρ_s and velocity $\mathbf{v}_s$. We again emphasize that the argument leading to equation (6) is a purely statistical one which considers quantum excitations of the *whole liquid*, and that nowhere is there any division of the helium itself into two real parts. Thus the name ' two-fluid model ' is to some extent misleading, as the ' fluids ' are only mathematical abstractions, but it seems to be here to stay. (This model has also been discussed by Dingle 1952 a, H. A. Kramers 1952, and Kronig 1953.)

We now proceed to derive two important equations which describe the motion of the two ' fluids '. The entropy of the helium arises solely from the thermal disorder and therefore will be associated only with the excitations. Transport of entropy will take place at the mean group velocity of the excitations and this may be shown to equal $\mathbf{v}_n$ (Feynman 1954), that is the entropy may be considered as moving with the normal fluid. Thus we can write down the following equation for the conservation of entropy :

$$\frac{\partial}{\partial t}(\rho S)+\operatorname{div}(\rho S\mathbf{v}_n)=0, \qquad \ldots\ldots(9)$$

S being the entropy of one gramme of liquid helium. This equation also implies that the motion of the fluids is reversible ; this turns out to be a good approximation although irreversible effects have to be considered later.

Finally we need an equation for the forces acting on the superfluid alone, and therefore write down the usual differential expression for the internal energy of a mass of liquid helium

$$dU = TdS - pdV + GdM$$

where G is the Gibbs energy of one gramme of liquid and dM is a change in the mass of the assembly. Let us now increase the mass of the assembly at constant volume by introducing particles which contribute only to the ground state, that is to the superfluid, so that $dS = 0$. As $dV = 0$, we may write $dU = GdM$; the sole result of the operation has been to increase the mass of superfluid by dM, therefore the potential energy of unit mass of superfluid must be G. The equation of motion of the superfluid is thus

$$\frac{\partial \mathbf{v}_s}{\partial t} = -\operatorname{grad} G. \qquad \ldots\ldots(10)$$

Collecting together equations (9) and (10), and adding the usual hydrodynamical equations for $\mathbf{j}$, the total momentum of the liquid, we obtain

$$\frac{\partial \rho}{\partial t}+\operatorname{div}\mathbf{j}=0 \qquad \ldots\ldots(11.1)$$

$$\frac{d\mathbf{j}}{dt}+\nabla p=0 \qquad \ldots\ldots(11.2)$$

$$\frac{\partial(\rho S)}{\partial t}+\rho S\operatorname{div}\mathbf{v}_n=0 \qquad \ldots\ldots(11.3)$$

$$\frac{d\mathbf{v}_s}{dt}+\nabla G=0. \qquad \ldots\ldots(11.4)$$

Having derived these equations, first given by Landau, we are now in a position to discuss the applications of the two-fluid model.

§ 5. APPLICATIONS OF THE TWO-FLUID MODEL

5.1. *Superfluidity*

Superfluidity was first observed in experiments in which the flow of liquid helium through fine slits and capillaries was seen to take place with virtually a complete absence of friction. Kapitza (1941 a) deduced that if the coefficient of viscosity had a finite value at all, it must be less than 10^{-11} poise, while Allen and Misener (1939) showed that the flow did not obey the ordinary equations of fluid motion. (Actually Allen and Misener's results for slits varying from 10^{-5} to 10^{-3} cm in width were rather complex, but in the narrowest slits the flow appeared independent of the pressure head.) This behaviour is referred to as superfluidity, and can in principle be understood from what has been said already, although some of the detail is still obscure.

Consider the flow of a liquid through a capillary. In general it is necessary to apply a pressure to maintain the flow, and most of the work done by this pressure appears as heat in the liquid. Without going into any details of the viscous processes, we can say that the resistance to flow arises because of a conversion of some of the directed kinetic energy of the liquid into disordered thermal motion. Or again, the relative motion of liquid and capillary sets up additional thermal motion in the liquid, and this is the source of the viscous reaction. Thus to maintain a flow of liquid helium at absolute zero through a tube, work must be done to provide the energy for the creation of any thermal excitations (phonons or rotons) which are produced. However, by a simple argument based on the transformation of coordinates, Landau (1941) showed that excitations can only be formed in this way if the velocity of the fluid through the capillary exceeds a certain critical value

$$\begin{aligned} v_c &= c \simeq 240 \text{ metres sec}^{-1} \text{ (phonons)} \\ v_c^2 \sqrt{(2\Delta/\mu)} &\simeq 60 \text{ metres sec}^{-1} \text{ (rotons)} \end{aligned} \qquad \ldots\ldots(12)$$

where Δ, μ and c have the same meanings as in equations (1) and (3). Thus if the velocity of the liquid is less than 60 metres sec^{-1} no excitations should be formed, and we reach the surprising conclusion that there will be no viscosity. (Actually it turns out that the values for the 'critical velocities' given in equations (12) are much too high, but we ignore this for a moment.) Hence the liquid at absolute zero should flow through a tube under the action of a vanishingly small pressure gradient at a rate equal to the critical velocity; if it flows more slowly it will be accelerated by the pressure gradient, if more quickly it will be retarded by viscous forces.

The flow of helium at higher temperatures is more complicated. At absolute zero the helium contains no excitations and is entirely superfluid; thus the discussion of the last paragraph shows that the superfluid has no viscosity: hence its name. At higher temperatures, the total motion of the helium must be resolved into that of the two-component (mathematical) fluids. As before, the motion of the superfluid will be free of viscous forces because of the difficulty of producing excitations in it. However, excitations already exist in the normal fluid and these, by interactions with the walls of the tube, can take up energy without restriction. Thus viscous effects should be associated with the motion of the normal fluid, but will be difficult to observe in flow experiments because they will be masked by the great volume of superfluid flow.

There appears to be little doubt that it is the difficulty of creating thermal excitations in the liquid which is responsible for the appearance of superfluidity. However, the mechanism whereby they are excited is obviously not that discussed by Landau, because the value of the critical velocity is found to be not of the order of 60 m sec^{-1} but only of the order of a few cm sec^{-1}, and is also a complicated function of temperature and slit width. This particular problem is one of the least understood aspects of liquid helium and is discussed later, in the section on turbulence. For the present we confine our attention to the region where there appears to be true superfluidity, that is where the velocities are sufficiently low not to produce excitations in the superfluid.

5.2. *The Normal Viscosity and Andronikashvili's Experiment*

As we have just discussed, measurements of the rate of flow of helium through a capillary do not enable us to determine the viscosity associated with the normal fluid. This 'normal viscosity' η_n is best found by measuring the viscous force on a body moving through the liquid; in an experiment of this type the superfluid exerts no force on the moving body and the total reaction is that due to the normal viscosity alone. The most straightforward method uses a rotation viscometer and the results of detailed measurements by Heikkila and Hollis-Hallett (1955) are shown in figure 6. We defer a discussion of the form of these results to § 6.1, only noticing here that the viscosity is of the same order as that of a gas.

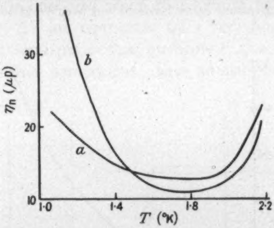


Figure 6. The 'normal' viscosity of liquid helium as determined (*a*) directly with a rotation viscometer, and (*b*) with an oscillating-disc viscometer plus a determination of the normal density (*after* Heikkila and Hollis-Hallett 1955, and Hollis-Hallett 1952).

One would suppose that the same value for the viscosity would be obtained by making measurements with an oscillating-disc viscometer, but the damping on such a disc is found to decrease steadily as the temperature falls, as is shown in figure 7 (Keesom and MacWood 1938). Before discussing this apparent discrepancy it is necessary to consider the important experiment of Andronikashvili (1948). This author mounted fifty thin metal discs very close together on a common axis, oscillated them about a vertical axis in a bath of liquid helium, and found that below the lambda point the period of the oscillating system decreased with falling temperature. This curious result is accounted for by the fact that while the superfluid is not affected by the motion of the discs, the normal viscosity is sufficient to drag round with the discs all the normal fluid in the narrow spaces between them. Thus the rotating system consists of the pile of discs plus the normal fluid accompanying them; as the density of the normal fluid varies with temperature, so will the moment of inertia and period of the system. Thus from the experimental observations one can associate a density with the normal fluid, and from what was said in § 4 it would seem that the density so determined should be the density of the normal fluid ρ_n. In fact the values of ρ_n found in this way (figure 8) are in good agreement with those calculated from equation (6) and also with values derived from the velocity of temperature waves which are discussed in § 5.4.

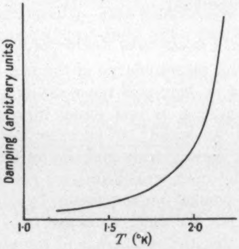


Figure 7. The damping on a disc oscillating in liquid helium, in arbitrary units as a function of temperature (*after* Keesom and MacWood 1938).

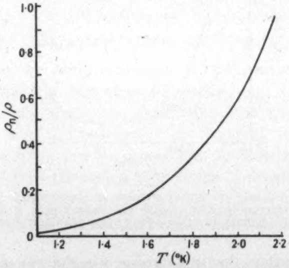


Figure 8. The ' normal ' density of liquid helium ρ_n (*after* Peshkov 1946).

It can now be seen why the damping of an oscillating disc decreases steadily with falling temperature whereas the viscosity does not. The damping of the disc is essentially due to viscous waves which the disc propagates in the liquid, and this damping is proportional to the product of the density and viscosity of the liquid. While for all other liquids the density is constant and the damping of the disc proportional to the viscosity alone, this is not so for helium. Clearly the density in question is not the total density of the helium, as the superfluid is not affected by the motion of the disc, and the position has been analysed in detail by Landau (1944). He showed that the motion of the helium is separable into two independent parts, one part corresponding to a potential flow of superfluid, while the other concerns only the normal fluid and is similar to the solution of the Navier–Stokes equation for a liquid of density ρ_n and viscosity η_n. Thus, in the absence of pressure and temperature gradients, the equation of motion of the normal fluid may be written :

$$\rho_n \frac{d\mathbf{v}_n}{dt} = \eta_n \nabla^2 \mathbf{v}_n.$$

Hence one ultimately obtains values for the viscosity which are in tolerable agreement with those obtained from the rotation viscometer, although, as is seen in figure 6, there is some discrepancy at temperatures below 1·4°K where the position is still somewhat obscure. It may be that the higher values for the normal viscosity given by the oscillation method are due to experimental errors which become more serious at low temperatures, and one possible source of error has recently been reported by Kolm and Herlin (1956). However, it has also been suggested by Ginsburg (1955) that the apparent viscosity is actually greater in the case of the oscillatory motion, because the non-equilibrium conditions are responsible for some form of turbulence in the liquid, as discussed in § 7. Further experiments seem to be required.

5.3. *The Fountain and Mechano-caloric Effects*

The fountain effect is observed when two helium baths at somewhat different temperatures T and $T+\Delta T$ are connected by a very fine capillary, as in figure 9. The capillary acts as a semi-permeable membrane, through which the superfluid may pass but not the normal component ; thus an osmotic pressure Δp is developed corresponding to the excess pressure of normal fluid in the hotter reservoir. This effect was first observed by Allen and Jones (1938), who used a slightly different experimental arrangement and produced such big temperature differences that the pressure flung the liquid up as a jet, hence the name fountain effect. The magnitude of the effect may be derived most simply by noting that in equilibrium there will be no acceleration of the fluids and therefore, by equation (11.4), grad $G = 0$. Using the

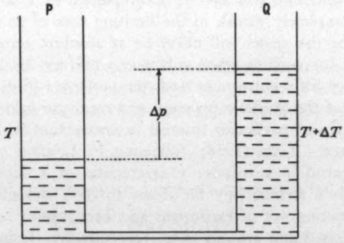


Figure 9. The fountain effect. The vessels P and Q are joined by a fine capillary ; if the liquid helium in Q is hotter than that in P an osmotic pressure Δp is developed between them.

general thermodynamic relation $dG = Vdp - SdT$, one thus obtains the relation

$$\frac{\Delta p}{\Delta T} = \rho S \qquad \ldots\ldots(13)$$

first given by H. London (1939) using a slightly different argument. Alternatively it is possible to calculate the pressure in terms of the constants of the excitations (see, for example, Ward and Wilks 1952). Measurements of the fountain effect are somewhat difficult to carry out, as large differences in the levels are produced by only very small temperature differences. There have been several sets of measurements between about 1°K and the lambda point (for example, Kapitza 1941 a, Meyer and Mellink 1947), and although there is considerable scatter on the results, it seems that H. London's relation is obeyed at least to within the accuracy of the measurements.

We now describe the inverse or mechano-caloric effect (Daunt and Mendelssohn 1939 a) by considering an isolated vessel with a very fine slit at the bottom. Superfluid will flow out through the slit, and as it has no thermal energy it carries away no entropy; thus the entropy per unit mass of liquid left in the vessel increases, and the liquid becomes warmer. Alternatively, if the vessel is held in an isothermal bath at temperature T, then a quantity of heat Q, the heat of transport, must be removed from the vessel as unit mass of superfluid flows out. If the process is reversible, as we assume, then Q is equal to TS; this relation has been confirmed to at least a few per cent, any discrepancies being probably within the experimental error (Chandrasekhar and Mendelssohn 1955, Brewer, Edwards and Mendelssohn 1955). Finally we note that as the superfluid leaving the vessel has zero entropy, one might suppose that here is a method of reaching absolute zero and violating the Third Law of thermodynamics. However, even in the narrowest slits, the flow of superfluid will also be accompanied by a flow of the normal fluid, which will only completely vanish in the limiting case of an infinitely narrow slit; thus helium leaving the vessel will never be at absolute zero. Nor is the process particularly useful for cooling other substances to very low temperatures because the helium has very little entropy at such temperatures (Simon 1950).

Measurements of the fountain pressure and mechano-caloric effect in the region of 1°K and below are of particular interest in connection with an earlier two-fluid theory due to Tisza. Tisza (1938), following F. London (1938), suggested that liquid helium behaved in a manner characteristic of a degenerate Bose–Einstein gas; using Landau's terminology he seems to have thought of the normal and superfluids as consisting not of *excitations* and background but of *atoms* or groups of atoms in the excited and ground states respectively. Using this picture he was able to account qualitatively for the effects already discussed and also to predict the existence of temperature waves (§ 5.5). Following naturally from this viewpoint, he supposed that the atoms in the superfluid ground state took part in normal Debye type lattice vibrations; in this case the term S in equation (13) does not refer to the whole entropy of the liquid but only to that part in excess of the entropy of the lattice vibrations (H. London 1939). In other words, on Tisza's model the phonons do not contribute to the fountain pressure. At 1°K the phonons are responsible for about 50% of the total entropy of the helium, at 0·8°K for about 80%, and at 0·6°K for almost 100%; thus it is possible to differentiate sharply between the theories of Landau and Tisza. Peshkov (1954) has made measurements down to 0·8°K and Bots and Gorter (1956) down to 0·2°K, and it appears that at these temperatures the fountain pressure is given fairly closely by equation (13) with S equal to the total entropy of the liquid. Thus these results, like those concerning temperature waves to be discussed later, confirm Landau's approach in which the normal fluid consists of *excitations* rather than excited *atoms*.

5.4. *Heat Transport*

Transport of heat takes place so readily in liquid helium that impracticably large heat inputs would have to be used to maintain measurable temperature gradients in the bulk liquid, and therefore nearly all the work so far has been done on helium in narrow slits or capillaries. In general the heat flow is not proportional to the cross-sectional area of the liquid; in the narrower channels and for not too large temperature gradients it does vary as dT/dx, but in wider channels it is approximately proportional to $(dT/dx)^{1/3}$. At present we consider the region where the flow depends linearly on the temperature gradient and where the magnitude of the flow may in fact be deduced from Landau's model; we return to the question of the heat flow in wider channels in § 7.1.

Consider two reservoirs of helium at temperatures T and $T+\Delta T$, connected by a very fine slit as in figure 9, so that virtually the whole fountain pressure $\Delta p = \rho S \Delta T$ is developed across it. As the slit is not infinitely narrow (as we assumed when considering the fountain effect) there will be some flow of normal fluid through it, under the action of the osmotic pressure difference Δp, together with a frictionless return flow of superfluid to conserve mass. The magnitude of these two counter current flows will be limited only by the viscosity of the normal component, the velocity of the normal fluid being given by Poiseuille's formula

$$v_n = \frac{B}{\eta_n}\frac{\Delta p}{\Delta x} \qquad \ldots\ldots(14)$$

where Δx is the length of the slit, and B is a constant depending on its geometry. Assuming as before that the relative motion of the two fluids takes place reversibly, the heat flow per unit area, ω, due to the transport of entropy by the normal fluid, will be given by

$$\omega = \rho S T v_n. \qquad \ldots\ldots(15)$$

Substituting (14) in (15) together with London's expression for Δp, we get

$$\omega = \frac{B(\rho T S)^2}{\eta_n}\frac{\Delta T}{\Delta x}. \qquad \ldots\ldots(16)$$

This result has been given both by Andronikashvili (1949) and by F. London and Zilsel (1948), who introduced the name 'internal convection' to describe this mechanism of heat transfer.

As mentioned above, this linear relation between the heat flow and the temperature gradient is only true for not too great heat flows, and in narrow slits of width less than about 100μ (see, for example, figure 10). As the constant B contains the width of the slit to the third power, the experimental verification of equation (16) is by no means easy. Hence it is not surprising that the agreement with some of the earlier measurements is not too close; however, recent work by Winkel, Broese van Groenou and Gorter (1955) seems to confirm that this relation is valid over quite a large region. They show that at spot temperatures the heat flow is proportional to the cube of the slit width as demanded by Poiseuille's equation (figure 11), and that values may be deduced for the viscosity of the normal fluid which are in reasonable agreement with more direct measurements, bearing in mind the accuracy of the experiments.

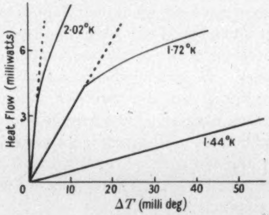


Figure 10. The heat flow through liquid helium in a slit of width $2{\cdot}4\mu$ as a function of the temperature difference across the slit. The relation is linear only for small heat flows at not too high temperatures (*after* Winkel, Groenou and Gorter 1955).

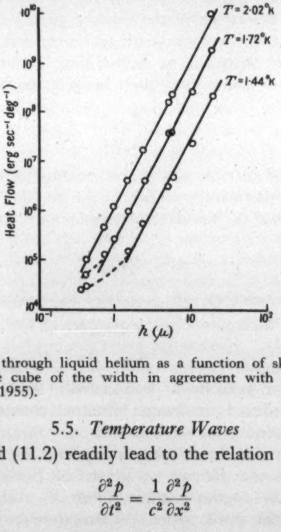


Figure 11. The heat flow through liquid helium as a function of slit width h. The flow varies approximately as the cube of the width in agreement with equation (16) (*after* Winkel, Groenou and Gorter 1955).

5.5. *Temperature Waves*

Equations (11.1) and (11.2) readily lead to the relation

$$\frac{\partial^2 p}{\partial t^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial x^2}$$

the usual equation for the propagation of density or pressure changes in a liquid.

However, there is the additional possibility, with the two-fluid model, of a fluctuation in the relative densities of the normal and superfluids while the total density remains constant ($\mathbf{j} = 0$). As a change in the relative densities of normal and superfluids is equivalent to a change in temperature, an oscillation of the two fluids relative to each other corresponds to an entirely new phenomenon, a true oscillation of temperature. Remembering that the thermal expansion coefficient of helium is so small that the adiabatic and isothermal compressibilities are nearly equal, one can solve the four equations of the model (11.1 to 11.4) and obtain the ordinary sound equation as before plus a second relation (Landau 1941)

$$\frac{\partial^2 S}{\partial t^2} = \frac{\rho_n}{\rho_s}\frac{C}{TS^2}\cdot\frac{\partial^2 S}{\partial x^2}$$

C being the specific heat of unit mass of helium. This is a wave equation for the propagation of variations of entropy or, what amounts to the same thing, of temperature variations ; thus the model predicts the existence of harmonic temperature waves with velocity

$$c_2 = \left(\frac{\rho_s}{\rho_n}\frac{TS^2}{C}\right)^{1/2}.$$

This expression for the velocity, first given by Landau, was confirmed by Peshkov (1946), who excited waves in the liquid by periodic heating ; he detected their presence with a thermometer and was able to tune them to resonance by setting up standing waves. Other measurements were subsequently performed (e.g. Pellam 1949) which showed that, in accordance with the wave equation, pulses of heat propagate through the liquid as compact pulses. Both types of experiment give the same values for the wave velocity, and these (figure 12) together with the known values of T, S and C lead to values of ρ_n/ρ_s which are in good agreement with those determined by other methods.

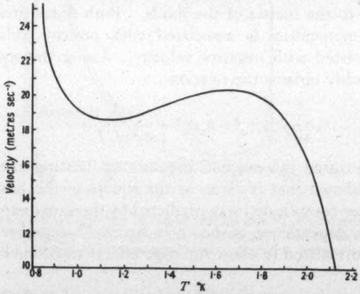


Figure 12. The velocity of temperature waves in liquid helium as a function of temperature (*after* Maurer and Herlin 1951).

Historically, it should be mentioned that the existence of temperature waves was first predicted by Tisza (1938) on the basis of a somewhat different model. However, at the time of writing his first paper, Landau was apparently unaware of this work, and seems to have thought that the velocity c_2 would be associated with sound of the usual type but with another velocity—hence he gave it the name 'second sound'. It was only after a failure (see Peshkov 1946) to excite such waves with a piezoelectric crystal that Lifshitz (1944) showed that a much more effective technique would be to generate a periodically varying temperature.

At temperatures appreciably below 1°K the position becomes more complex and, as is discussed in § 6.3, it now appears that in general a wave-like propagation of temperature variation is not possible below about 0·7°K. However, it is important to note that resonance techniques (Peshkov 1952) as well as pulse measurements confirm the existence of temperature waves down to at least 0·85°K, and that the velocity rises rapidly as the temperature falls, in agreement with Landau's formula. In contrast, Tisza's theory, which is based on fundamentally different concepts (§ 5.3), predicts that the velocity should fall in this region.

5.6. *The Momentum of Heat Flow*

As discussed in the last sections, heat flow in liquid helium is accompanied by motion of the two fluids; therefore as the fluids have momentum we arrive at the unusual result that a flow of heat has momentum associated with it. This momentum was first observed by Kapitza (1941 b), who showed that a flow of heat impinging on a vane mounted on a torsion wire gave rise to a pressure which deflected the vane. However, Kapitza's experiment did not show the dependence of the pressure on temperature which one would now calculate, probably because of the rather uncertain geometry of his apparatus. The momentum flux associated with each fluid is $\rho\mathbf{v}.\mathbf{v}$; thus in a linear flow of heat there is a reaction pressure $\rho_s v_s^2 + \rho_n v_n^2$ on the source, due to the inertia of the fluids. Both fluids give positive contributions, as positive momentum is associated with positive velocity, and negative momentum is associated with negative velocity. Using the equations of the two-fluid theory one readily obtains the relation

$$\rho_s v_s^2 + \rho_n v_n^2 = \frac{\omega^2}{c_2^2 \rho C T}. \qquad \ldots\ldots(17)$$

Hall (1954) has measured the reaction pressure on heating elements immersed in liquid helium and shown that it varies as the square of the heat flow, and as temperature in the rather complicated way predicted by the above expression. The form of this temperature dependence is shown in figure 13, together with experimental points arbitrarily normalized to allow for edge effects associated with the finite size of the heaters.

Another ingenious demonstration of the existence of the relative motion of the two fluids in a temperature wave has been given by Pellam and Morse (1950) and Pellam and Hanson (1952). As the experiment involves the use of a modified Rayleigh disc in liquid helium, we briefly recall that, in general, a disc in a moving fluid tends to align itself crosswise to the direction of the particle velocity. The torque acting on the disc is proportional to the kinetic energy of the motion, that is to ρv^2, where ρ is the density of the fluid and v the particle velocity. Because the magnitude of the torque is quadratic in the particle velocity, there will be a resultant mean torque on the disc even though the velocity varies harmonically and is directed in opposite directions for successive half-cycles. Now the motion of the two fluids in a temperature wave in liquid helium has kinetic energy $\frac{1}{2}\rho_n v_n^2$ associated with the normal fluid and $\frac{1}{2}\rho_s v_s^2$ associated with the superfluid. Both these terms give rise to torques which will be additive, even though it follows from equation (8) that the particle velocities in the two fluids must be in antiphase if we assume that there are no density fluctuations ($\mathbf{j} = 0$). The total torque may be readily related to the heat current in the temperature wave, and calculated values for a given disc are shown in figure 14; the experimental results of Pellam and Hanson (1952) are in satisfactory agreement.

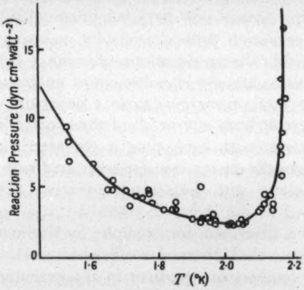


Figure 13. The momentum of heat flow exemplified by the reaction pressure on a source of heat in liquid helium. The full curve gives the theoretical expression of equation (17), and the points the normalized experimental values (*after* Hall 1954).

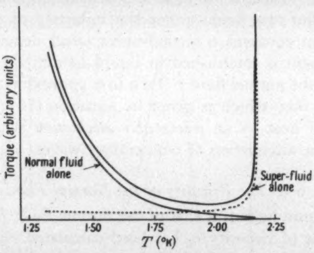


Figure 14. The torque on a Rayleigh disc in liquid helium due to temperature waves, as a function of temperature. The upper curve shows the total torque; the lower curves give the torques associated with the motion of the normal and superfluids respectively (*after* Pellam and Hanson 1952).

§ 6. THE EXCITATION GAS

In the preceding applications of the two-fluid model the normal fluid has been regarded as an abstract continuous medium, but many of the results could have been obtained by considering the fluid as a gas-like assembly of excitations. Thus the fountain pressure is the excess pressure of the excitation gas in the hotter vessel, and a flow of heat down a temperature gradient corresponds to a flow of excitations maintained by the fountain pressure. In this section we consider more complex parameters of the liquid, such as the coefficients of viscosity and sound absorption, whose magnitudes can only be derived from a consideration of the excitations themselves. Both these coefficients characterize processes in which the distribution function of the excitations is disturbed from its equilibrium value, so that a flow of energy is set up tending to restore the equilibrium. Such behaviour has been examined in a very general way by Khalatnikov (1952 a), who considers the most general form of excitation defined solely by its energy ϵ and momentum p. Khalatnikov takes the four linear equations of motion (11.1) to (11.4) together with the Landau expression for the Hamiltonian of an excitation ($H = \epsilon(\mathbf{p}) + \mathbf{p} \cdot \mathbf{v}_s$) and then calculates the flow of energy due to a departure from equilibrium. In general this flow may result from any or all of three possible types of disturbance, and Khalatnikov associates with each type a characteristic 'kinetic coefficient' α, β or γ. α characterizes the energy transfer associated with the 'normal' viscosity and β the transfer associated with the second viscosity. (This second viscosity is sometimes also described as the dynamic viscosity; it is associated with density changes in a system, as is described, for example, by Karim and Rosenhead (1952).) The third coefficient γ is the 'thermal coefficient' and concerns the transfer of energy when there is a temperature gradient in the excitation gas, but although it is somewhat analogous to the thermal conductivity of an ordinary gas the analogy has to be used cautiously. Thus γ has no immediate relation to the flow of heat in liquid helium itself, which as we have seen is brought about by the relative and *reversible* motion of the two fluids under the influence of the fountain pressure. The thermal coefficient concerns a second-order effect depending on the fact that if a temperature gradient is established in liquid helium there will also be a concentration gradient in the normal fluid. To a first approximation this will not affect the value of the heat flow which is given by equation (16); however, it does give rise to a small flow of heat by an *irreversible mechanism* which is of consequence when one considers the attenuation of temperature waves.

6.1. *The Viscosity of the Normal Fluid*

The viscosity of liquid helium is that of the normal fluid, that is of the excitation gas, and the coefficient of viscosity η_n has been calculated on this basis by Landau and Khalatnikov (1949) using standard kinetic theory methods. Below about 1·8°K, where the excitation gas is not too dense and the interactions not too strong, the excitations are assigned scattering cross sections and mean free paths; then by setting up the transport equation an expression is obtained for the momentum transfer and hence the viscosity. The calculations are somewhat lengthy because collisions may occur between phonons and phonons, phonons and rotons, and rotons and rotons; also the results for each type of collision must be averaged over the whole of momentum space. It is not possible here to go into any of the details and we merely indicate the outline of the treatment and the results obtained.

As both the phonons and the rotons transport momentum, there is both a phonon and a roton contribution to the viscosity. For lack of more detailed information a roton is treated as a sort of particle; interactions between rotons lead to a viscosity which is independent of temperature but of indeterminate magnitude. (This independence of temperature is quite analogous to the fact that the viscosity of a perfect gas is independent of its density.) In the case of the phonons it is possible to obtain an exact expression for the viscosity, as their interaction energies are related to the curvature of the phonon part of the excitation spectrum. (As may be seen from figure 5 (*b*), this spectrum is only given by the linear relation $\epsilon = pc$ for values of p close to zero.) The phonon viscosity is limited principally by the scattering of the phonons by the heavier rotons, and therefore increases with falling temperature because of the rapidly decreasing number of rotons. The calculated value is in quite good agreement with the observed rise in the normal viscosity at temperatures below 1·4°K. However, the close numerical agreement which is obtained is probably somewhat fortuitous because many of the required data (e.g. $\partial\Delta/\partial\rho$, $\partial^2\Delta/\partial\rho^2$) have to be deduced or estimated from other experiments rather indirectly, as for example is done by Atkins (1955); Landau and Khalatnikov themselves remark that no great accuracy can be expected. Also there is still some uncertainty over the experimental value of the viscosity at low temperatures as mentioned in § 5.1. However, the theory gives a good semi-quantitative account of the behaviour of the viscosity.

6.2. *The Attenuation of Pressure and Temperature Waves*

The coefficient of the absorption of sound or pressure waves for a frequency of 14 Mc/s is shown in figure 15. Above about 3°K the absorption is accounted for by the classical expression for the attenuation due to viscosity and thermal conduction; the high value between 2° and 3°K is associated with the lambda transition, and has been discussed, in detail, by Pippard (1951). For the present, however, we are concerned with the region below the lambda point where the behaviour is obviously quite different from that above, and where the absorption has been calculated by Khalatnikov (1950, 1952 b) on the basis of the excitation gas model. The passage of a pressure wave disturbs the equilibrium of the helium, and therefore of the excitation gas; then, as is well known, there are absorption effects except in the two limiting cases when equilibrium is restored either very quickly or very slowly with respect to the period of the sound waves. The absorption arises by virtue of the transfer of energy in non-equilibrium conditions

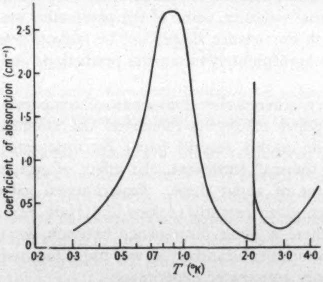


Figure 15. The absorption of 14 Mc/s sound waves in liquid helium as a function of temperature (*after* Pellam and Squire 1947, Chase 1953, Newell and Wilks 1956).

by the three mechanisms discussed in § 6; by solving the appropriate equations for the wave motion Khalatnikov obtains expressions for the attenuation of sound waves in terms of the coefficients of first and second viscosity and of the thermal coefficient.

In his first paper Khalatnikov (1950) pointed out that the viscosity of the normal fluid can only account for about one-fifth of the attenuation observed by Pellam and Squire (1947). As the coefficient of expansion of helium is small, any temperature changes due to the passage of the sound wave will also be small and therefore Khalatnikov postulated that the excess absorption was due to second viscosity rather than to the thermal coefficient. The calculation of the magnitude of the second viscosity requires a knowledge of two relaxation times, characteristic of the approach to equilibrium of an excitation gas in which the number of phonons and the number of rotons have non-equilibrium values. In principle the relaxation times can be calculated, but in practice this is very complicated; therefore Khalatnikov deduced their temperature dependence and then obtained their absolute magnitudes by fitting his expression for the attenuation to the experimental values. Having obtained these two parameters, he was able to predict from Pellam and Squires' results, which only extended down to 1·6°K, that there should be a maximum in the absorption at about 0·9°K and that it should be accompanied by dispersion (Khalatnikov 1950). Quite some time afterwards, both the maximum and the dispersion were observed and found to be approximately of the predicted magnitude (Chase 1953, Chase and Herlin 1955, Newell and Wilks 1956); at the lowest temperatures, however, the absorption is much greater than was expected.† Finally, we should mention that the calculations are very involved and that, as in the calculation of the normal viscosity, some of the parameters are only known approximately; thus too much importance should not be laid on detailed numerical agreement. However, the agreement between the predictions and what was afterwards discovered is striking.

Khalatnikov also considered the attenuation of temperature waves in the bulk liquid. In his first paper (1950) he calculated the attenuation due to first and second viscosity, while in his second paper he took into account the process characterized by the thermal coefficient, the effect of which was to increase the attenuation by a factor of about three. Experimental values of the absorption down to 1°K have been given recently (Atkins and Hart 1954, Pellam and Hanson 1954) and although there is some discrepancy between the two sets of results, it seems that the attenuation is substantially as was predicted, rising as the temperature falls. Here again theory anticipated experiment.

6.3. *Mean Free Path Effects*

Although the normal fluid is an assembly of discrete excitations, equations (11.1) to (11.4) implicitly assume that the two fluids are continuous (mathematical) media. We now consider what modifications to the results already obtained are involved when the structure of the normal fluid is taken into account, and it is useful to begin by considering the propagation of temperature waves. Suppose we ignore the superfluid for the moment and consider only the excitation gas, which is the normal fluid. Then it may be shown (Ward and Wilks 1951, 1952, Dingle 1952b) that, provided the conservation laws of energy and momentum are valid for interactions between the excitations, a periodic disturbance of the distribution function will propagate itself as a wave motion. But temperature variations may be regarded as particular kinds of changes in the distribution function, and therefore the excitation gas will propagate temperature waves. Dingle also shows that the velocity of the waves so arrived at is given by $c_2^2 = (\rho/\rho_n)\,TS^2/C$ which approximates to the Landau expression at low temperatures, when ρ_n is small and $\rho \simeq \rho_s$.† At higher temperatures the deviations from the Landau formula are considerable because it is no longer permissible to ignore the motion of the superfluid. The point is that fluctuations in the gas of excitations correspond to changes in the normal density ρ_n, and as the ordinary density of the helium is assumed to remain constant, the relation $\rho_n + \rho_s = \rho$ implies a drift of superfluid associated with the temperature fluctuations. At low temperatures where the density fluctuations are small this drift may be neglected, but at higher temperatures it is essential to take it into account as is done in the two-fluid model.

---

† It now seems likely that the excess absorption is associated with temperature differences set up by the wave, which may no longer be ignored at the lowest temperatures (Dransfeld, Newell and Wilks, to be published.)

From the above discussion it is clear that an essential criterion for the propagation of second sound is that the mean free path of excitations between interactions should be small compared with the wavelength of the second sound. Now the work of Landau and Khalatnikov, discussed above, shows that the collisions which propagate the second sound are those considered when calculating the viscosity. From their work on the viscosity it is known that the mean free paths between collisions increase very rapidly with falling temperature, so that by about 0·6° they are of the order of 1 cm. This figure is considerably greater than the wavelength of the second sound in any experiments so far performed and one would therefore expect the propagation of temperature waves to break down by this temperature (Gorter 1952, Atkins 1953), as indeed is shown by several experiments (for example, those of Kramers, Peski-Tinbergen, Wiebes, van den Burg and Gorter 1954). Figure 16 (*a*) shows the shape of a temperature pulse received at 0·9°K as viewed on an oscillograph trace triggered by the transmitted pulse; the pulse is quite sharp and is of about the same width as the transmitted signal. Figure 16 (*b*) shows the corresponding received signal at about 0·2°K, viewed on the same time scale; the pulse arrives much sooner, as would be expected from the calculated velocity of temperature waves at this temperature, but it has become so very broad that a wave equation of propagation is hardly applicable. As all these experiments

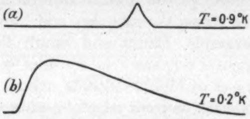


Figure 16. The shape of a heat pulse transmitted through liquid helium and viewed on an oscilloscope trace, as a function of temperature; at 0·9°K the pulse is quite compact and characteristic of a wave-like form of propagation (*after* Atkins and Osborne 1950).

---

† The values of ρ_s at 1·0 and 1·2°K are about 0·99ρ and 0·97ρ respectively, so that even at 1·2°K this formula gives the velocity correct to 1½%.

have been conducted in comparatively narrow cylinders of helium, the mean free paths of the excitations have also been greater than the diameter of the container. Thus the flow of heat in the helium should be very similar to the flow in a dielectric medium whose thermal conductivity is limited only by boundary scattering. Ziman (1954) has worked out in detail the change of pulse shape which would be expected on such a model, and obtains results very similar to those observed by Atkins and Osborne (1950) (figure 16).

Similar mean free path effects are to be expected in the behaviour of the viscosity; thus η_n should fall below the Landau and Khalatnikov value when the mean free paths characteristic of the viscous processes approach the dimensions of the experimental apparatus. This effect should be most marked in narrow slits at low temperatures but very few data are yet available. However, the results of Winkel, Broese van Groenou and Gorter (1955) give an anomalously high value of the heat flow in an 0·4μ slit at the lowest temperatures, and it may well be that such effects are limiting the viscosity, the mean free path calculated by Landau and Khalatnikov being of the order of the slit width at this temperature.

§ 7. TURBULENCE

As we mentioned when discussing the derivation of ρ_n, the equations of motion for the liquid are valid only to the first order in $v_s - v_n$, or what amounts to the same thing in v_s or v_n; thus we might expect them to be inadequate for experiments involving high velocities of the fluids. Several authors have discussed the form that the second-order terms should take, and there is very little agreement between them (see, for example, Daunt and Smith 1954). However, the only experiment in which it appears necessary to consider quadratic terms is the propagation of temperature waves of high amplitude, where shock effects are observed (Osborne 1951); either the back or front edge of a pulse becomes steeper according to the temperature (figure 17). Khalatnikov (1952 c) has shown that this behaviour is consistent with equations of motion quadratic in v^2 given by Landau (1941), and there seems little need to investigate this particular question further. It is generally found that before the velocities of flow become sufficiently large to call for the use of these quadratic terms, other processes occur which drastically modify the behaviour of the helium. These processes, which we characterize quite loosely by the name turbulence, are by no means fully understood, so we first present the experimental facts and then go on to mention possible interpretations.

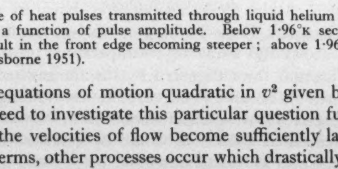


Figure 17. The shape of heat pulses transmitted through liquid helium and viewed on an oscilloscope trace, as a function of pulse amplitude. Below 1·96°K second-order effects, at high amplitudes, result in the front edge becoming steeper; above 1·96°K the back edge becomes steeper (*after* Osborne 1951).

7.1. *Viscosity and Heat Flow Experiments*

Kapitza (1941 b) observed that the heat flow through an annulus of helium between a thin tube and a narrow rod was greatly reduced if the rod was rotated at speed. This was the first observation of the now generally recognized behaviour of the liquid, that when the velocities of the constituent fluids exceed ' critical ' values of the order of a few cm sec^{-1}, its properties differ considerably from those outlined above. The transition to the perturbed state is well marked, as is illustrated by measurements of the viscosity and heat flow. Hollis-Hallett (1950, 1955) has shown that the damping of an oscillating disc rises steeply at a certain amplitude of swing, in a manner which suggests the onset of new dissipative forces entirely absent at lower amplitudes (figure 18). Similarly, Hung, Hunt and Winkel (1952) showed that if the temperature gradient in a fine slit exceeds a certain value, then the heat flow is no longer governed by equation (16), but assumes lower values and even becomes proportional to the cube root of the temperature gradient. Lacking a better name, we will refer to this condition characterized by higher dissipative forces as one of turbulence, although without committing ourselves as to its precise nature.

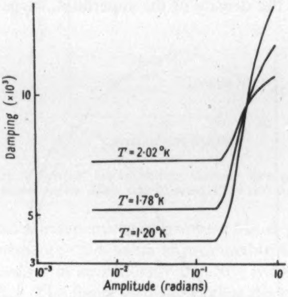


Figure 18. The damping on a disc oscillating in liquid helium as a function of amplitude (*after* Hollis-Hallett 1955).

A valuable step in analysing these experiments was taken by Gorter and Mellink (1949) who postulated that the equations of motion of the two-fluid model would still be valid in the turbulent region provided that additional frictional terms were included. Besides the viscous forces η_n acting on the normal fluid alone, they postulated a mutual friction force acting between the two fluids and of the form $F_{sn} = A\rho_s\rho_n(v_s - v_n)^3$, the cube relation being chosen to give the correct dependence of heat flow on temperature gradient. Hence they write

$$\rho_s \frac{d\mathbf{v}_s}{dt} = -\frac{\rho_s}{\rho}\operatorname{grad} p + \rho_s S \operatorname{grad} T - A\rho_s\rho_n(\mathbf{v}_s - \mathbf{v}_n)^3,$$

$$\rho_n \frac{d\mathbf{v}_n}{dt} = -\frac{\rho_n}{\rho}\operatorname{grad} p - \rho_s S \operatorname{grad} T + A\rho_s\rho_n(\mathbf{v}_s - \mathbf{v}_n)^3 + \eta_n \nabla^2 \mathbf{v}_n, \quad \ldots\ldots(18)$$

these equations being equivalent to Landau's if $A = 0$ (Dingle 1949).

These equations account moderately well for the heat flow experiments, although the values of the constant A deduced from measurements made with different sized slits at different temperatures vary by over a factor of two. However, they do not seem capable of accounting for the flow of the liquid through narrow channels such as described by Hung, Hunt and Winkel (1952) and more particularly by Atkins (1951). As these flow experiments are both difficult to carry out and to interpret, too much weight should not perhaps be attached to some quantitative discrepancy, but even so the equations appear to be inadequate.

Equations (18) as they stand also fail to interpret the oscillating disc experiments of Hollis-Hallett (1952); for example, figure 19 shows the excess damping which would be expected using a value of mutual friction deduced from the heat flow measurements. Actually the excess damping observed varies with temperature in a rather similar manner to the density of the superfluid, suggesting the presence of a frictional force F_s which acts in the superfluid alone. The existence of such a superfluid viscous term in addition to the mutual friction force is confirmed by measurements of the viscosity using a rotation viscometer (Heikkila and Hollis-Hallett 1955). In the steady state the mutual friction term cannot give rise to any torque, but at the higher velocities a contribution to the torque is observed over and above that due to the viscosity of the normal fluid. Finally, to round off the argument, we mention that the flow experiments of Atkins (1952) would seem to indicate that if only a force F_s were present (and no mutual friction), then the heat transport that would be deduced from measurements of fluid flow would be grossly too large. Thus to account for the above results it seems that both terms F_s and F_{sn} must be added to the equations of motion. Of course there is no need to postulate two different mechanisms; if some form of viscosity is set up in the superfluid, one would expect it to be accompanied by some interaction with the normal component.

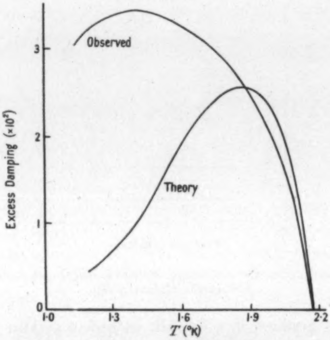


Figure 19. The excess damping on a pile of discs oscillating in liquid helium which results when large amplitudes of oscillation are used. (Amplitude and period of oscillation: 0·215 radian and 2·60 sec respectively.) The curves show the experimental values and the values predicted by the Gorter–Mellink theory (*after* Hollis-Hallett 1952).

7.2. *The Critical Velocity*

An important feature of turbulence is shown in the results of Allen and Misener (1939) for the velocity of flow in slits 10^{-4} to 10^{-5} cm wide (figure 20). The

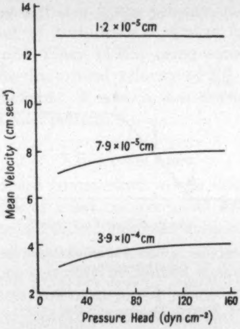


Figure 20. The velocity of flow of liquid helium through fine channels at 1·2°K. Channel width as shown (*after* Allen and Misener 1939 and Atkins 1952).

velocity has a value almost independent of pressure but depending on the particular conditions of the experiment; this value is presumably the critical velocity, above which the viscous forces set in so suddenly that any further increase of velocity is energetically impossible. The figure also shows that the rate of flow, and therefore the critical velocity, decreases by a factor of about 4 for an increase in slit width of about 30 times. A dependence of critical velocity on the smallest linear dimension of the liquid is observed in many different types of experiments and appears to be one of the most fundamental aspects of the onset of turbulence. Thus Atkins (1951) has observed the flow of helium through capillaries up to 10^{-2} cm diameter and shown that the critical velocity continues to decrease with increasing diameter up to at least 10^{-2} cm, although in this case the flow of the normal fluid becomes so considerable that it is difficult to obtain accurate values. The experiments with oscillating disc and rotation viscometers, referred to above, give a clear indication of the onset of turbulence; and in the case of the rotation viscometer it seems reasonable to assume that the critical dimension is the gap between the two cylinders.

Several recent experiments have helped to clarify the position by showing how the mutual friction depends on the velocities of the two fluids. Winkel, Delsing and Gorter (1955) have measured the fountain effect across slits of such width that the full fountain pressure is not developed because the slit is not completely impermeable to the normal fluid. From the equations of motion they show that the reduction in fountain pressure in their experiments is directly proportional to the constant of mutual friction, and are therefore able to obtain values for this constant. Winkel (1955) has shown that if helium is caused to flow from vessel P to Q through a narrow slit by the supply of heat to Q, then the motion does not cease immediately on cutting off the supply of heat. Observations of this 'overshoot' effect by Winkel, Delsing and Poll (1955) lead to further values for the mutual friction and also to accurate values of the critical velocity. Finally, Winkel, Broese van Groenou and Gorter (1955) have made and analysed new measurements of heat flows in narrow slits at sufficiently low heat current densities so as

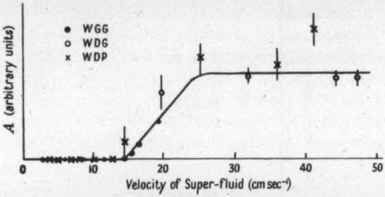


Figure 21. The coefficient of mutual friction A as a function of the velocity of the superfluid v_s (in arbitrary units).
WGG : Winkel, Groenou and Gorter (1955) ;
WDG : Winkel, Delsing and Gorter (1955) ;
WDP : Winkel, Delsing and Poll (1955).
(*After* Winkel, Groenou and Gorter 1955.)

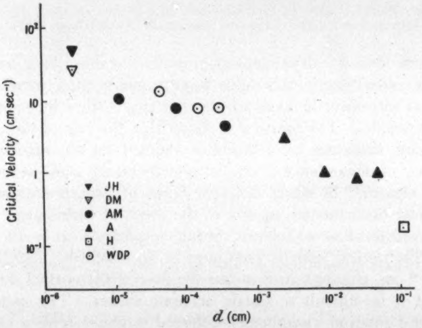


Figure 22. The critical velocity of the superfluid as a function of the smallest linear dimension of the helium d.
JH : from film thickness (Jackson and Henshaw 1950) ;
DM : from film thickness (Daunt and Mendelssohn 1939) ;
AM : from flow (Allen and Misener 1939) ;
A : from flow (Atkins 1951) ;
H : from viscosity (Hollis-Hallett 1952) ;
WDP : from 'overshoot effect' (Winkel, Delsing and Poll 1955).
(*After* Dash 1954 and Winkel, Delsing and Poll 1955.)

to be able to observe the transition to turbulence. From all these data it is possible to decide which is the crucial velocity, that of the normal or of the superfluid or their relative velocity. Figure 21, taken from Winkel, Broese van Groenou and Gorter, shows values of the mutual friction constant A, obtained by the different Leyden methods, for a given temperature and slit width, plotted against the velocity of the superfluid. We see that in all the rather different experiments there seems to be a unique relation between the values of A and the superfluid velocity. Moreover this would not be the case if the normal or relative velocities were taken as abscissae, for in flow type experiments $V_n \simeq V_s$, but in heat transport experiments $V_n \simeq V_s \rho_s/\rho_n$. Thus it appears that it is the velocity of the superfluid which gives the criterion for the onset of turbulence; similar conclusions have also been reached by Dash (1954) and Vinen (1955) using somewhat different arguments. Figure 22 summarizes how the critical velocity of the superfluid depends on the linear dimensions of the helium; it is seen that the results from many different experiments fall into a common pattern.

7.3. *Vortex Lines*

The experimental evidence summarized in the last section suggests that at a certain velocity, of the order of a few cm sec^{-1}, and which depends on the dimensions of the apparatus, the superfluid becomes subject to frictional forces. This implies that some forms of excitation are being created in the superfluid, but, on Landau's theory, excitations can only be formed if the velocity of flow is at least 60 metres sec^{-1} for rotons and 240 metres sec^{-1} for phonons (§ 5.1). Clearly some other mechanism has yet to be taken into account and we now consider what may have been omitted in Landau's approach.

The crux of the matter seems to be that the argument used by Landau to deduce values for the critical velocity implicitly assumes that the flow of the liquid is irrotational potential flow. (It assumes that when an excitation is created the velocity of the whole mass of moving liquid is reduced by the same small amount.) This point was well appreciated by Landau and he had already shown that the flow was potential ($\mathrm{rot}\, v_s = 0$) by arguments based on his quantum hydrodynamics. However, we have previously mentioned objections to the hydrodynamics, and these have recently been confirmed by an experiment which shows that on occasion the flow is definitely not potential. Hall (1955) has measured the torque needed to accelerate a cylindrical vessel containing liquid helium to a constant angular velocity, and has thus shown that when a steady state has been reached the whole mass of the helium is rotating. At first sight this result might seem to be at variance with Andronikashvili's experiment (§ 5.2) in which velocities of the same order of magnitude were involved but only the normal fluid took part in the motion. However, the two experiments are not inconsistent, for in Hall's experiment it took about a minute for the superfluid to be accelerated to a steady velocity, and the acceleration was disproportionately small at the beginning of the motion. Thus one would hardly expect the superfluid to be set in motion at all in Andronikashvili's experiment which concerned oscillatory motion of period about a second. It is this time effect which also accounts for the fact that temperature waves are not affected by turbulence, even though the velocities of the fluids may exceed the critical value for uni-directional flow.

Given time it appears that the whole mass of liquid helium can be brought into rotation, and it now becomes necessary to consider what configuration the rotating helium is likely to adopt. It has been realized for some years (H. London 1946, F. London 1954, Landau and Lifshitz 1955) that the most energetically favourable state would be one in which cylindrical annuli of the liquid performed irrotational rotations about the axis, the annuli being separated by vortex sheets whose separation decreased with increasing angular velocity. Onsager (1949) has postulated that the vortex sheets consist of arrays of vortex lines and that these will be quantized. This concept has been discussed in detail by Feynman (1955) but as yet only in a rather qualitative way. However, although much is yet to be analysed, it seems clear that such vortices could be a source of mutual friction, and, being quantized, would presumably only be created under certain critical conditions.

7.4. *Rotating Helium*

The clearest evidence for the existence of the vortex lines discussed in the last section is an experiment on the propagation of temperature waves, by Hall and Vinen (1955). These authors mounted a temperature wave resonator, complete with heater and thermometer, in a vessel of liquid helium and rotated the whole assembly at constant velocity. When a steady state had been reached, it was found that there was an additional attenuation of the waves which was proportional to the angular velocity of the system. The absorption increased if the cryostat was disturbed by shaking or if the rotating system was accelerated, but the most significant result was the presence of additional absorption just due to steady rotation. A rather similar experiment has also been performed by Wheeler, Blakewood and Lane (1955), who observed a greater attenuation probably because the geometry of their apparatus was such as to induce more turbulence in the liquid.

Thus there is good ground for supposing that vortex lines are the key to the behaviour of liquid helium in the super-critical condition, but several problems have still to be elucidated. There is still no really satisfactory treatment relating the critical velocities discussed in § 7.2 with the dimensions of the apparatus, although Feynman (1955) has made a start on this problem, and shown how the critical velocity could decrease with increasing slit width. There is also no explanation at all as to why the heat flow in the super-critical region varies approximately as the cube root of the temperature gradient. Again it is not entirely clear how the vortices are related to the two-fluid model, although they probably should be considered as another form of excitation with a very large scattering cross section. As the total energy per unit volume of the vortex lines is small compared with the heat energy (Feynman 1955), they will probably make little contribution to the density of the normal fluid. This view seems to be confirmed by two recent experiments on rotating helium which involve the density of the normal fluid: Hall and Vinen (1955) found that the velocity of temperature waves is unaffected by rotation, and Andronikashvili and Kaverkin (1955) have shown that the fountain pressure is also unaffected.

§ 8. THE HELIUM FILM

One of the most characteristic features of helium II is the phenomenon of film flow first described by Rollin (1936) and Rollin and Simon (1939), and investigated in an elegant series of experiments by Daunt and Mendelssohn (1939 b). The essence of the experimental arrangement is shown in figure 23; if the lower part of an empty beaker is lowered into a bath of helium, helium from the bath creeps up over the rim of the beaker until the liquid levels inside and outside are equal.

Similarly, if a beaker containing liquid is placed above the bath, the helium creeps over the rim and returns to the lower level in the bath. To account for this behaviour one recalls that if a solid surface is in contact with any liquid having a positive angle of contact, a thin film of liquid forms on it due to van der Waals forces. In the case of helium this film will be of the order of 10^{-6} cm thick, and below the lambda point the superfluid component of the liquid in the film will be fully mobile, so that mass flow can take place under extremely small potential gradients. (In the beaker experiment this potential is provided by the variation of the free energy of the vapour with height above the lower liquid level.) Although this brief outline sums up the essential behaviour of the film, there is still much uncertainty over the details of the picture, principally because of the difficulties of establishing the experimental facts, and some of these points are now considered in more detail.

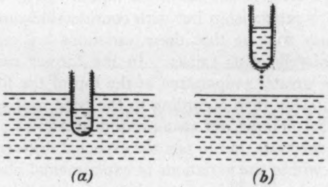


Figure 23. (*a*) If an empty beaker is lowered into a bath of helium II, film flow occurs over the surface of the beaker until the levels are equalized. (*b*) If the beaker is now lifted above the bath, the helium flows out of the beaker via the film, drops off the bottom of the beaker and returns to the bath (*after* Daunt and Mendelssohn 1939).

8.1. *The Thickness of the Film*

The thickness of the film below the lambda point was first measured by Daunt and Mendelssohn (1939 b) and Kikoin and Lazarew (1938), who determined the volume of liquid in the film covering a known large area. These authors obtained a value of about 3×10^{-6} cm, essentially independent of temperature, and this result has since been confirmed several times. All the earlier experiments also led to the conclusion that above the lambda point the thickness of the film became quite small, about 10^{-7} cm, a result apparently inconsistent with the suggestion of Frenkel (1940) and Schiff (1941) that the film owes its existence to van der Waals forces, which will presumably show little change in passing through the lambda point. However, both Frenkel and Schiff pointed out that although a film of helium II would be able to establish itself very quickly because of its high mobility, the same would not be true of helium I. In fact, it seems that great precautions have to be observed in order to obtain the true equilibrium thickness of the film; any extraneous influx of heat may evaporate the film at a rate greater than it can be replaced. Moreover, as Meyer (1954) pointed out, a very small reduction in pressure to just below the saturation value will lead to a very large reduction in the thickness of the film, and the excess helium will drop off very rapidly. However, on restoring the pressure to its original value, the film above the lambda point is only able to build up to its original thickness very slowly; it is not superfluid and its low thermal conductivity will prevent it discharging the high heat of condensation which must accompany the formation of liquid from the vapour phase. Thus it is not perhaps surprising that there are considerable discrepancies between different observations of the film. As mentioned above, nearly all the earlier work suggested that the film became quite thin above the lambda point; but the recent work of Ham and Jackson (1954) has shown that, by taking sufficient precautions against heat influxes and temperature fluctuations, a thick film is also observed above the lambda point. Thus although the position is still confused, it seems likely that under carefully controlled equilibrium conditions the film is of at least comparable thickness above and below the lambda point, and this is confirmed by measurements of the absorption isotherms (see, for example, Long and Meyer 1953).

Finally, according to Schiff and Frenkel, if the van der Waals forces obey the usual inverse sixth power law, then the thickness of the film should vary as $h^{-1/3}$, where h is the height above the surface of the liquid. Most workers have obtained results closer to an $h^{-1/2}$ relationship but with considerable variations of the absolute magnitude, and it may well be that these variations are due to extraneous heat influxes and other miscellaneous causes. In the former case the incoming heat would probably cause greater evaporation at the top of the film, and the thickness would fall off more quickly than according to the cube root relation (Meyer 1955). Schiff and Frenkel also pointed out that the film thickness should depend on the surface energy of the substrate, but in fact the expected difference even in favourable cases turns out to be within the variations in experimental observations (Smith and Boorse 1955). It must be added, however, that there is still no satisfactory explanation of the report by Ham and Jackson (1954) that if the substrate is contaminated with solid air the film may be considerably thicker, a result in accord with the earlier observations of Bowers and Mendelssohn (1950) that the flow rate in such cases is greatly augmented.

8.2. *Transport Properties*

Below the lambda point the film appears to be sufficiently thick for the properties of the helium in the film to be essentially those of the bulk liquid, and the film exhibits superfluidity and permits of a high heat transport. This superfluidity was studied in detail by Daunt and Mendelssohn (1939 b), who found that the rate of flow was proportional to the perimeter of the vessel; their results are shown in figure 24. The rate of mass transport through the film will be proportional to the velocity of flow, to the thickness of the film and to the fraction of the helium free to move, that is to the superfluid density. As discussed in § 7.2, one would expect the velocity of flow to be equal to the critical velocity which will in turn depend on the thickness of the film. Figure 24 shows that the rate of flow varies with temperature in much the same way as the density of the superfluid does; this implies that the product of film thickness and critical velocity is independent of temperature. Therefore, as the film thickness is approximately independent of temperature at least below the lambda point, the critical velocity appears to be a function of film thickness which is at least approximately independent of temperature, a result which is not in disagreement with the conclusions of § 7.3. (It should be mentioned that as the thickness of the film varies with height there is the very considerable difficulty that it is not really clear which value of the thickness is to be taken as characteristic of the critical velocity.) We also note that both Ambler and Kurti (1952) and Waring (1955) have reported a considerable increase in the rate of film transport below about 0·6°K. This result does not fit into the general picture, and it would be interesting to confirm it by experiments under rather more isothermal conditions than have hitherto been possible.

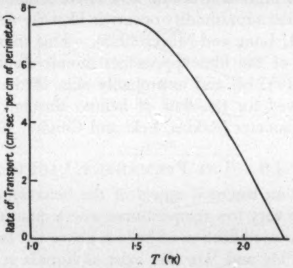


Figure 24. The rate of transport of helium II through the film, as a function of temperature (*after* Daunt and Mendelssohn).

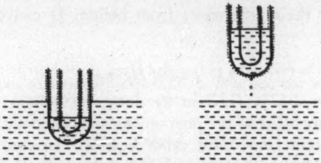


Figure 25. The double-beaker experiment. The helium in the beakers flows through the film to the lowest gravitational level ; note that the flow takes place between the inner and outer beakers even though any difference in the levels is imperceptible (*after* Daunt and Mendelssohn 1946).

One of the most significant of the film flow experiments is that of Daunt and Mendelssohn (1946) which is illustrated in figure 25 ; on lifting the double beaker assembly from the helium bath, the liquid from both inner and outer beakers ran out into the bath. In particular flow took place from the inner to outer beaker, although no pressure or temperature gradient could be observed between them. Likewise, on lowering the beaker into the liquid, a flow occurred similarly but in the opposite direction. This behaviour implies that the kinetic energy of the moving liquid in the film is not dissipated on entering the bulk liquid but must be converted into free energy, that is to say the flow of the liquid must be purely potential flow.

8.3. *The Unsaturated Film*

A comparatively large number of experiments have been made on the unsaturated film, that is to say film in equilibrium with unsaturated helium vapour and therefore not so thick as the saturated film. The results of these experiments are not easy to interpret, for apparently quite similar experiments do not always give the same results. At present it hardly seems possible to correlate all the observations into a coherent system, and it seems that some unaccounted factors may be influencing at least some of the experiments. The position has been reviewed by Long and Meyer (1953) ; here we only mention that the lambda anomaly in the specific heat is smeared out in thin films and occurs at a lower temperature (Frederikse 1949) ; the temperature at which superfluidity occurs is likewise depressed (Bowers, Brewer and Mendelssohn 1951, Long and Meyer 1955). This shift in the lambda temperature as the thickness of the film approaches atomic dimensions has been briefly discussed by Ziman (1953 b), and is probably akin to the depression of the onset of superfluidity observed for the flow of helium through narrow channels of the order of 5×10^{-7} cm diameter (Atkins, Seki and Condon 1956).

§ 9. THE PERMANENT LIQUIDS

We have seen that an essential aspect of the behaviour of helium II is that it remains liquid down to very low temperatures where quantum effects are important. However, helium I (the modification of helium above the lambda point), the isotope ^{3}He, and mixtures of ^{3}He and ^{4}He also exist as liquids at these temperatures and have very different properties from helium II. Thus any satisfactory theory of helium II must indicate whether or not these other ‘ permanent ’ liquids behave like helium II. Therefore we now give a short account of their properties, and then discuss whether their differences from helium II can be predicted from the present theory.

9.1. *Liquid Helium I*

Liquid ^{4}He above 2·17°K (helium I) does not exhibit superfluidity, a high transport of heat or any other of the phenomena associated with the two-fluid model, and is not very different from most other liquids. In contrast to helium II the specific heat has a value which only varies slightly with temperature (Keesom and Clusius 1932), and the expansion coefficient is positive. Viscous processes and heat flow may be described in the usual way by coefficients of viscosity (de Troyer, van Itterbeek and van den Berg 1951) and thermal conductivity (Bowers 1952), and the absorption of sound is entirely accounted for by the classical expression involving these two coefficients. As in the case of helium II, the large influence of the zero-point energy results in the molar volume being very much greater than one would expect from classical theory. The atoms are comparatively free in this large volume and the behaviour of helium I is in some ways more like that of a gas than a liquid ; for instance, the coefficient of viscosity is small and *decreases* with falling temperature. However, the experiments of Tjerkstra (1952) show that if the helium is compressed to higher densities, thereby reducing the space available for each atom, the viscosity eventually rises with falling temperature as in other liquids (figure 26). The thermal conductivity also has a small value which decreases with falling temperature, and is related reasonably well to the coefficient of viscosity by the usual gas kinetic expression $k = 2{\cdot}5\eta$ (Grenier 1951). Virtually all the differences between helium I and helium II may be accounted for by the fact that helium I does not exhibit any of the phenomena characteristic of the two-fluid model, and it is satisfactory to see that this result is inherent in the theory of helium II. It is an essential feature of the two-fluid model that the normal density ρ_n is less than the total density ρ, and also that ρ_n increases steadily with temperature, thus at some temperature ρ_n must become equal to ρ; there is then no superfluid component and the whole concept of the two-fluid model becomes invalid. Indeed, the lambda point may be characterized by $(\rho_n)_\lambda = \rho$.

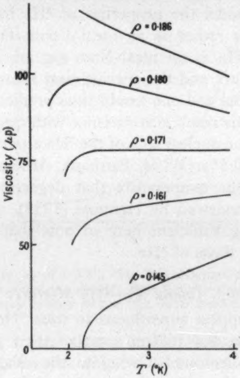


Figure 26. The coefficient of viscosity of liquid helium I as a function of density. At low densities the viscosity decreases with falling temperature as in a gas, but at higher densities a more liquid-like behaviour is observed (*after* Tjerksta 1953).

9.2. *Liquid* ^{3}He

Liquid ^{3}He was first observed by Sydoriak, Grilly and Hammel (1949), and found to have a critical point at 3·34°K in agreement with the value previously predicted by de Boer and Lunbeck (1948) using a reduced quantum-mechanical equation of state. The density of the liquid is very low (about 0·08 g cm^{-3} at 1°K compared with 0·14 g cm^{-3} for ^{4}He), indicating the large role played by the zero-point energy. As in the case of ^{4}He, the liquid does not solidify under its vapour pressure, at least down to 0·5°K (Osborne, Abraham and Weinstock 1952), and a pressure of at least 30 atmospheres must be applied to obtain the solid phase. The adsorption isotherms (Hoffman, Edeskuty and Hammel 1956) are very like those for ^{4}He. Thus it appears that the interatomic forces in liquid ^{3}He and ^{4}He are quite similar and give rise in both cases to liquids with a very open structure which remain fluid down to the lowest temperatures.

At present the number of experiments reported on ^{3}He is small compared with the very large number on ^{4}He, but it is already clear that the lighter isotope has quite different properties. Measurements of the specific heat and entropy give no indication of any lambda or other type of transition down to 0·3°K (Osborne, Abraham and Weinstock 1954, Roberts and Sydoriak 1954, de Vries and Daunt 1954). Weinstock, Osborne and Abraham (1949) using a flow-type experiment have shown that the viscosity has a quite normal value of about 25 μp and that there is no evidence of superfluidity down to 1·05°K. Daunt and Heer (1950) showed that the addition of ^{3}He to ^{4}He depresses the lambda point as defined by the onset of superfluid flow in the film; a solution containing 89% of ^{3}He has a lambda temperature of 0·38° and it appears likely that pure ^{3}He will not give rise to a superfluid film at all. Thus it seems well established that liquid ^{3}He does not exhibit the phenomena characteristic of the two-fluid model.

For lack of a better model the properties of ^{3}He have been discussed on the assumption that it behaves rather as an ideal Fermi–Dirac gas, even though the limitations on regarding ^{4}He as an ideal Bose gas are known to be considerable (§ 9.4). The vapour pressure and the specific heat measurements are in quite fair agreement with such a model and one would thus predict a degeneracy temperature of about 5°K. However, this result is at variance with measurements of the magnetic susceptibility arising from the nuclear spin of the ^{3}He atoms; this susceptibility obeys Curie's law down to about 0·5° K (W. M. Fairbank, Ard and Walters 1954) indicating that it is only at about this temperature that degeneracy sets in. The present position, which has been reviewed by Hammel (1955), is obscure and will not be considered further, it being sufficient here to note that the properties of ^{3}He are substantially different from those of ^{4}He.

9.3. *Liquid* ^{3}He–^{4}He *Mixtures*

At first sight it might appear superfluous to treat ^{3}He–^{4}He mixtures as distinct forms of permanent liquids, but in fact some of their properties differ markedly from those of either pure isotope. Owing to the relative scarcity of ^{3}He most experiments have been made with solutions containing only a small percentage of this isotope, but as we are concerned principally with the properties of ^{4}He this is of no great consequence. At low concentrations, the ^{3}He atoms will be virtually independent of each other and will not exhibit any of the bulk properties of ^{3}He, they rather act as a solute or impurity (Landau and Pomeranchuk 1948). Thus they give rise to an osmotic pressure whose magnitude is determined by the law of van't Hoff, and which may be developed between two vessels joined by a very fine slit which is permeable only to superfluid ^{4}He (Taconis, Beenakker and Dokoupil 1950). According to the nature of the two-fluid components of helium II, it is to be expected that solute atoms will interact only with the excitations, and will therefore be carried along with the normal fluid (Landau and Pomeranchuk 1948, Khalatnikov 1952 d). Thus in the presence of a temperature gradient all the ^{3}He is carried to the cold part of the apparatus by the flow of normal fluid, and this is the basis of the heat flush method of concentrating ^{3}He from ^{3}He–^{4}He mixtures (Lane, H. A. Fairbank, Aldrich and Nier 1948). The association of the ^{3}He with the normal fluid is confirmed by Andronikashvili disc measurements in ^{3}He–^{4}He mixtures (Pellam 1955); the density of the normal fluid does not tend towards zero at low temperatures but appears to take up a constant value equal to about twice the mass of the ^{3}He atoms present in unit volume (figure 27). This increase in ρ_n indicates that the ^{3}He atoms are dragged round with the discs; that the magnitude of the increase is about twice the mass density of the ^{3}He atom may be accounted for by the fact that the moving atoms have to push aside ^{4}He atoms and hence have a greater inertial resistance than their own mass (Feynman 1954). Pellam also obtained semi-qualitative values for the viscosity of the normal component, and found that the value of η_n no longer increased below 1·4°K as in pure ^{4}He. The viscosity of the pure liquid rises because the decreasing number of rotons results in a longer mean free path for the phonons and a correspondingly greater transport of momentum. The additional ^{3}He atoms act as a comparatively large number of scattering centres, the mean free path remains constant, and the viscosity decreases steadily with falling temperature as the number of phonons decreases.

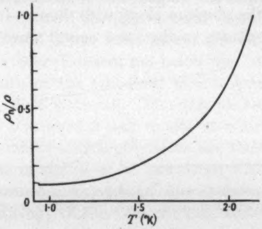


Figure 27. The normal density ρ_n of a 3·3% ^{3}He–^{4}He solution, as measured by Andronikashvili's method. At low temperatures the density does not tend towards zero but takes up a value of the order of the effective density of the ^{3}He (*after* Pellam 1955).

At first sight it might appear that a small percentage of ^{3}He atoms in liquid ^{4}He would form scattering centres which would considerably reduce the transport of heat in the liquid. However, these impurity atoms are not fixed in space as they would be in a solid, but are free to move with the normal fluid. Thus as the high heat transport in liquid ^{4}He is due to the relative motion of the two fluids, it will not be much affected if some ^{3}He atoms are carried along with the normal fluid. This is well shown by the experiments of Lynton and H. A. Fairbank (1950), who observed that a pulse of heat is not appreciably attenuated by the addition of small amounts of ^{3}He. It is to be noted that the above discussion assumes that the ^{3}He atoms are distributed uniformly through the liquid, but in fact they will tend to diffuse to the colder parts of the apparatus and form a concentration gradient there. In the experiments of Lynton and Fairbank, a uniform concentration of ^{3}He was ensured by the use of single pulses in order to keep the total heat flow very low, but if an attempt is made to measure heat flows *under steady uni-directional conditions*, then a concentration gradient of ^{3}He will be set up. In measurements of this type, the concentration gradient near the colder surface will prevent the free movement of ^{3}He atoms, and forces will be set up tending to restrain the motion of the normal fluid. Thus the measured heat flow is reduced from the very high value typical of ^{4}He to a quite moderate value (Beenakker, Taconis, Lynton, Dokoupil and van Soest 1952). Finally, as shown in figure 28, the addition of a small amount of ^{3}He (say 1%) to liquid ^{4}He will permit the liquid to transmit sharp pulses of temperature waves down to at least 0·2°K (King and H. A. Fairbank 1954). As discussed in § 6.3, temperature waves are not observed in pure ^{4}He below about 0·7°K because there are so few interactions between the excitations that equilibrium in the normal fluid cannot be established in distances of the order of a wavelength. However, even for a 1% concentration, the ^{3}He atoms are quite close together and, as is shown by the heat flush experiments, they interact with the excitations ; thus equilibrium may be established in the normal fluid and temperature waves observed. King and Fairbank also showed that the presence of ^{3}He atoms has a large effect on the velocity of the waves (figure 29), because the ^{3}He changes the values of both ρ_n and the specific heat and entropy of the liquid (Pomeranchuk 1949). At very low temperatures the normal fluid consists almost entirely of ^{3}He atoms and a temperature wave is essentially a variation in the density of these atoms which drag the phonons along with them. In this region the velocity of the waves should approximate to that of a sound wave in a gas of ^{3}He atoms,

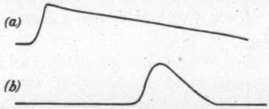


Figure 28. The shape of heat pulses through liquid helium and viewed on an oscilloscope trace : (*a*) pure ^{4}He at 0·45°K, (*b*) ^{4}He containing 0·32% ^{3}He at 0·45°K (*after* King and H. A. Fairbank 1954).

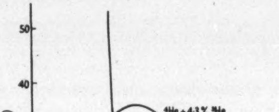


Figure 29. The velocity of temperature waves in ^{3}He–^{4}He solutions as a function of temperature (*after* King and H. A. Fairbank 1954).

and therefore be of the order of the velocity of the ^{3}He atoms, that is $(3kT/\mu)^{1/2}$ where μ is the effective mass (Feynman 1954). A more exact expression has been given by Pomeranchuk, namely $(5kT/3\mu)^{1/2}$, and although this does not appear to be in quantitative agreement with the measured values it is at least qualitatively correct.

As mentioned previously, the lambda point is depressed by the addition of ^{3}He ; this may be accounted for by the fact that the ^{3}He atoms increase the effective density of normal fluid, and therefore the criterion for the lambda point $[(\rho_n)_\lambda = \rho]$ is attained at lower temperatures. Measurements have also been made on the specific heat of mixtures, on the heat of mixing, and on the relative concentration of the two isotopes in the gaseous and liquid phases. These last experiments have been reviewed in detail (Daunt 1952, Beenakker and Taconis 1955), but do not at present seem to throw much light on the properties of pure ^{4}He.

9.4. *Statistics and the Lambda Point*

We have seen that dilute solutions of ^{3}He in helium II behave in a manner which may be predicted by an extension of the theory of helium II ; however, the same is not true of ^{3}He itself, and it seems that this is due to the influence of statistics. Thus ^{3}He presumably obeys Fermi statistics, whereas the theory of ^{4}He assumes that the particles obey Bose statistics, and that permutations of the atoms leave the wave function quite unchanged. It is significant that if one calculates the quantity ρ_n of equation (6) (the normal density of the two-fluid model) for Bose–Einstein and Fermi–Dirac ideal gases, one finds that ρ_n is equal to the total density except for a Bose–Einstein gas below the ' condensation ' temperature. Thus the essential feature of the two-fluid model appears to be a characteristic of an assembly obeying Bose statistics. We may also mention here that de Boer and Cohen (1955) have discussed a cell model in which each cell contains more than one atom, and in which exchange effects are therefore important. It is by no means clear how this model is to be associated with the present theory ; but for a cell model with four atoms in each cell the statistics are responsible for a very large difference in the entropy of the two isotopes, such as is observed (figure 30).

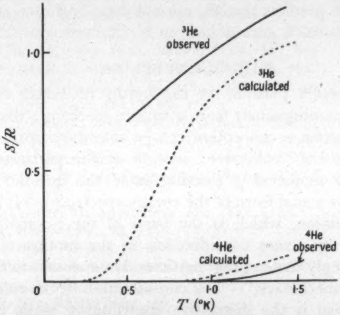


Figure 30. The entropy of ^{3}He and ^{4}He as calculated using a cell model with four atoms in each cell (*after* de Boer and Cohen 1955).

The importance of statistics was first stressed by F. London (1938), who pointed out that an ideal gas of the same density as liquid helium, and obeying Bose statistics, should exhibit an unusual type of phase transition at 3·14°K, with a specific heat anomaly somewhat similar to that observed in the liquid. He therefore suggested that the lambda transition was the analogy in the liquid of the ' condensation ' in the perfect gas, and on this basis Tisza was able to predict the existence of temperature waves. However, any such analogy cannot be a very close one, for liquid helium is not an ideal gas but an assembly of real atoms packed together as closely as they can get. From the specific heat and thermal conductivity measurements below 0·6°K, it is clear that there is a contribution to the thermal energy akin to that in a Debye type solid, and this certainly cannot arise in an ideal gas. The temperature of the lambda point *decreases* with increasing density of the liquid, in contrast to what would be expected for an ideal gas. The observed scattering cross section of helium II for neutrons (Sommers, Dash and Goldstein 1955) is quite different from that calculated using the gas model (Goldstein, Sweeney and Goldstein 1950), and the scattering of light is much less (Jakovlev 1943) than is predicted (Galanin 1940, Schiff 1941). Finally, as discussed in §§ 5.3 and 5.5, Tisza's theory leads to quite incorrect values for the fountain pressure and the velocity of temperature waves at low temperatures.

Although the ideal gas treatment is plainly inadequate, Feynman (1953 a) has shown that a form of Bose condensation is to be expected even in an assembly of strongly interacting particles. Feynman proceeds by writing down the complete partition function for the liquid, which he then simplifies and evaluates by approximate methods. Even the simplified function is difficult to evaluate, and Feynman concluded that the predicted transition was third order rather than second; however, Chester (1955) claims that a more accurate evaluation of the partition function leads to a second-order transition as is observed. It is typical of the difficulties in carrying out this evaluation that it is not entirely clear whether Feynman's partition function also leads to a liquid–gas transition of the normal type; see, for example, ter Haar (1954) and Chester (1955), who advance opposing views. We note also that Feynman predicts that the specific heat begins to rise in the helium I region as the lambda point is approached, as is observed.

§ 10. CONCLUSION

To sum up the present position we can hardly do better than quote Landau (1949): " It follows unambiguously from quantum mechanics that for every slightly excited macroscopic system a conception can be introduced of ' elementary excitations ', which describe the ' collective ' motion of the particles and which have certain energies ϵ and momenta p (leaving aside the question as to the actual dependence $\epsilon(p)$, i.e. the actual form of the energy spectrum). It is this assumption, indisputable in my opinion, which is the basis of the microscopical part of my theory. On the contrary, every consideration of the motion of individual atoms in the system of strongly interacting particles is in contradiction with the first principles of quantum mechanics." The consideration of the role of the *excitations* rather than of the *atoms* is the distinctive contribution made by Landau to the theory of helium II; the ' normal fluid ' is an aggregate of excitations rather than of excited atoms, as postulated by Tisza. It is for this reason that the phonons which are responsible for the behaviour of second sound and the fountain effect at low temperatures enter naturally into Landau's theory but not into Tisza's.

REFERENCES

ALLCOCK, G. R., and KUPER, C. G., 1955, *Proc. Roy. Soc.* A, **231**, 226.
ALLEN, J. F., and JONES, H., 1938, *Nature, Lond.*, **141**, 243.
ALLEN, J. F., and MISENER, A. D., 1939, *Proc. Roy. Soc.* A, **172**, 467.
AMBLER, E., and KURTI, N., 1952, *Phil. Mag.*, **43**, 260.
ANDRONIKASHVILI, E. L., 1948, *J. Exp. Theor. Phys.*, **18**, 424, 429; 1949, *Ibid.*, **19**, 535.
ANDRONIKASHVILI, E. L., and KAVERKIN, I. P., 1955, *J. Exp. Theor. Phys.*, **28**, 126.
ATKINS, K. R., 1952, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **1**, 169; 1953, *Phys. Rev.*, **89**, 526; 1955, *Ibid.*, **98**, 319.
ATKINS, K. R., and CHASE, C. E., 1951, *Proc. Phys. Soc.* A, **64**, 826.
ATKINS, K. R., and EDWARDS, M. H., 1955, *Phys. Rev.*, **97**, 1429.

Atkins, K. R., and Hart, K. H., 1954, *Canad. J. Phys.*, **32**, 381.

Atkins, K. R., and Osborne, D. V., 1950, *Phil. Mag.*, **41**, 1078.

Atkins, K. R., Seki, H., and Condon, E. U., 1956, *Phys. Rev.*, **102**, 582.

Atkins, K. R., and Stasior, R. A., 1953, *Canad. J. Phys.*, **31**, 1156.

Beenakker, J. J. M., and Taconis, K. W., 1955, *Progress in Low Temperature Physics* (Ed. C. J. Gorter) (Amsterdam: North-Holland), p. 108.

Beenakker, J. J. M., Taconis, K. W., Lynton, E. A., Dokoupil, Z., and van Soest, G., 1952, *Physica*, **18**, 433.

de Boer, J., and Cohen, E. G. D., 1955, *Physica*, **21**, 79.

de Boer, J., and Lunbeck, R. J., 1948, *Physica*, **14**, 510.

Bots, G. J. C., and Gorter, C. J., 1956, *Physica*, **22**, 503.

Bowers, R., 1952, *Proc. Phys. Soc.* A, **65**, 511.

Bowers, R., Brewer, D. F., and Mendelssohn, K. M., 1951, *Phil. Mag.*, **42**, 1445.

Bowers, R., and Mendelssohn, K. M., 1950, *Proc. Phys. Soc.* A, **63**, 1318.

Brewer, D. F., Edwards, D. O., and Mendelssohn, K. M., 1955, *Proc. Phys. Soc.* A, **68**, 939.

Casimir, H. B. G., 1938, *Physica*, **5**, 495.

Chandrasekhar, B. S., and Mendelssohn, K. M., 1955, *Proc. Phys. Soc.* A, **68**, 857.

Chase, C. E., 1953, *Proc. Roy. Soc.* A, **220**, 116.

Chase, C. E., and Herlin, M. A., 1955, *Phys. Rev.*, **97**, 1447.

Chester, G. V., 1955, *Phys. Rev.*, **100**, 455.

Dash, J. G., 1954, *Phys. Rev.*, **94**, 825.

Daunt, J. G., 1952, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **1**, 209.

Daunt, J. G., and Heer, C. V., 1950, *Phys. Rev.*, **79**, 46.

Daunt, J. G., and Mendelssohn, K. M., 1939 a, *Nature, Lond.*, **143**, 719; 1939 b, *Proc. Roy. Soc.* A, **170**, 423, 439; 1946, *Nature, Lond.*, **157**, 839.

Daunt, J. G., and Smith, R. S., 1954, *Rev. Mod. Phys.*, **26**, 172.

Dingle, R. B., 1949, *Proc. Phys. Soc.* A, **62**, 648; 1952 a, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **1**, 111; 1952 b, *Proc. Phys. Soc.* A, **65**, 374.

Fairbank, H. A., and Wilks, J., 1955, *Proc. Roy. Soc.* A, **231**, 545.

Fairbank, W. M., Ard, W. B., and Walters, G. K., 1954, *Phys. Rev.*, **95**, 566.

Feynman, R. P., 1953 a, *Phys. Rev.*, **91**, 1291; 1953 b, *Ibid.*, **91**, 1301; 1954, *Ibid.*, **94**, 262; 1955, *Progress in Low Temperature Physics* (Ed. C. J. Gorter) (Amsterdam: North-Holland), p. 17.

Feynman, R. P., and Cohen, M., 1956, *Phys. Rev.*, **102**, 1189.

Franchetti, S., 1954, *Nuovo Cim.*, **12**, 743.

Frederikse, H. P. R., 1949, *Physica*, **15**, 860.

Frenkel, J., 1936, *Wave Mechanics, Elementary Theory* (Oxford: University Press); 1940, *J. Phys., Moscow*, **2**, 365.

Galanin, A., 1940, *J. Exp. Theor. Phys.*, **10**, 1267.

Ginsburg, V. L., 1955, *J. Exp. Theor. Phys.*, **29**, 244, *Soviet Physics*, **2**, 170.

Goldstein, L., Sweeney, D., and Goldstein, M., 1950, *Phys. Rev.*, **77**, 319; **80**, 141.

Gorter, C. J., 1952, *Phys. Rev.*, **88**, 681.

Gorter, C. J., and Mellink, J. H., 1949, *Physica*, **15**, 285.

Grenier, C., 1951, *Phys. Rev.*, **83**, 598.

ter Haar, D., 1954, *Phys. Rev.*, **95**, 895.

Hall, H. E., 1954, *Proc. Phys. Soc.* A, **67**, 485; 1955, *Conférence de Physique des Basses Températures, Paris* (Paris: Institut International du Froid), p. 63.

Hall, H. E., and Vinen, W. F., 1955, *Conférence de Physique des Basses Températures, Paris* (Paris: Institut International du Froid), p. 66.

Ham, A. C., and Jackson, L. C., 1954, *Phil. Mag.*, **45**, 1084.

Hammel, F., 1955, *Progress in Low Temperature Physics* (Ed. C. J. Gorter) (Amsterdam: North-Holland), p. 78.

Heikkila, W. J., and Hollis-Hallett, A. C., 1955, *Canad. J. Phys.*, **33**, 420.

Henshaw, D. G., and Hurst, D. G., 1953, *Phys. Rev.*, **91**, 1222.

Hercus, G. R., and Wilks, J., 1955, *Phil. Mag.*, **45**, 1163.

Hoffman, C. J., Edeskuty, F. J., and Hammel, E. F., 1956, *J. Chem. Phys.*, **24**, 124.

HOLLIS-HALLETT, A. C., 1950, *Proc. Phys. Soc.* A, **63**, 1367 ; 1952, *Proc. Roy. Soc.* A, **210**, 404 ; 1955, *Progress in Low Temperature Physics* (Ed. C. J. Gorter) (Amsterdam : North-Holland), p. 64.

HUNG, C. S., HUNT, B., and WINKEL, P., 1952, *Physica*, **18**, 629.

JAKOVLEV, I. A., 1943, *J. Phys., Moscow*, **7**, 307.

KAPITZA, P. L., 1941 a, *J. Exp. Theor. Phys.*, **11**, 581 ; *J. Phys., Moscow*, **5**, 59 ; 1941 b, *J. Exp. Theor. Phys.*, **11**, 1 ; *J. Phys., Moscow*, **4**, 181.

KARIM, S. M., and ROSENHEAD, L., 1952, *Rev. Mod. Phys.*, **24**, 108.

KEESOM, W. H., and CLUSIUS, K., 1932, *Proc. Acad. Sci. Amst.*, **35**, 307, *Common. Leider.* No. 219e.

KEESOM, W. H., and MACWOOD, G. E., 1938, *Physica*, **5**, 737.

KEESOM, W. H., and TACONIS, K. W., 1938 a, *Physica*, **5**, 270 ; 1938 b, *Ibid.*, **5**, 161.

KHALATNIKOV, I. M., 1950, *J. Exp. Theor. Phys.*, **20**, 243 ; 1952 a, *Ibid.*, **23**, 8 ; 1952 b, *Ibid.*, **23**, 21 ; 1952 c, *Ibid.*, **23**, 253 ; 1952 d, *Ibid.*, **23**, 169.

KIKOIN, A. K., and LAZAREW, B. G., 1938, *Nature, Lond.*, **142**, 289.

KING, J. C., and FAIRBANK, H. A., 1954, *Phys. Rev.*, **93**, 21.

DE KLERK, D., HUDSON, R. P., and PELLAM, J. R., 1954, *Phys. Rev.*, **93**, 28.

KOLM, H. H., and HERLIN, M. A., 1956, *Phys. Rev.*, **102**, 607.

KRAMERS, H. A., 1952, *Physica*, **18**, 653.

KRAMERS, H. C., VAN PESKI-TINBERGEN, T., WIEBES, J., VAN DEN BURG, F. A. W., and GORTER, C. J., 1954, *Physica*, **20**, 743.

KRAMERS, H. C., WASSCHER, J. D., and GORTER, C. J., 1952, *Physica*, **18**, 329.

KRONIG, R., 1953, *Physica*, **19**, 535.

KRONIG, R., and THELLUNG, A., 1952, *Physica*, **18**, 749.

LANDAU, L. D., 1941, *J. Exp. Theor. Phys.*, **11**, 592 ; *J. Phys., Moscow*, **5**, 71 ; 1944, *J. Exp. Theor. Phys.*, **14**, 112 ; *J. Phys., Moscow*, **8**, 1 ; 1947, *Ibid.*, **11**, 91 ; 1949, *Phys. Rev.*, **75**, 884.

LANDAU, L. D., and KHALATNIKOV, I. M., 1949, *J. Exp. Theor. Phys.*, **19**, 637, 709.

LANDAU, L. D., and LIFSHITZ, E., 1955, *Dokl. Akad. Nauk., S.S.S.R.*, **100**, 669.

LANDAU, L. D., and POMERANCHUK, I., 1948, *Dokl. Akad. Nauk., S.S.S.R.*, **59**, 669.

LANE, C. T., FAIRBANK, H. A., ALDRICH, L. T., and NIER, A. O., 1948, *Phys. Rev.*, **73**, 256.

LIFSHITZ, E., 1944, *J. Exp. Theor. Phys.*, **14**, 116 ; *J. Phys., Moscow*, **8**, 110.

LOKKEN, J. E., 1954, *Canad. J. Phys.*, **32**, 359.

LONDON, F., 1938, *Nature, Lond.*, **141**, 643 ; 1945, *Rev. Mod. Phys.*, **17**, 310 ; 1954, *Superfluids*, Vol. 2 (New York : Wiley).

LONDON, F., and ZILSEL, P. R., 1948, *Phys. Rev.*, **74**, 1148.

LONDON, H., 1939, *Proc. Roy. Soc.* A, **171**, 484 ; 1946, *Proceedings of the Cambridge Low Temperature Conference* (London : Physical Society), p. 48.

LONG, E., and MEYER, L., 1953, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **2**, 1 ; 1955, *Phys. Rev.*, **98**, 1616.

LYNTON, E. A., and FAIRBANK, H. A., 1950, *Phys. Rev.*, **80**, 1043.

MAURER, R. D., and HERLIN, M. A., 1951, *Phys. Rev.*, **81**, 444.

MEYER, L., 1954, *Phys. Rev.*, **93**, 655 ; 1955, *Ibid.*, **97**, 22.

MEYER, L., and MELLINK, J. H., 1947, *Physica*, **13**, 197.

NEWELL, J. A., and WILKS, J., 1956, *Phil. Mag.*, **1**, 588.

ONSAGER, L., 1949, *Nuovo Cim.*, **6**, Suppl. **2**, 249.

OSBORNE, D. V., 1951, *Proc. Phys. Soc.* A, **64**, 114.

OSBORNE, D. W., ABRAHAM, B. M., and WEINSTOCK, B., 1951, *Phys. Rev.*, **82**, 263 ; 1952, *Ibid.*, **85**, 158 ; 1954, *Phys. Rev.*, **94**, 202.

PELLAM, J. R., 1949, *Phys. Rev.*, **75**, 1183 ; 1955, *Conférence de Physique des Basses Températures, Paris* (Paris : Institut International du Froid), p. 24.

PELLAM, J. R., and HANSON, W. B., 1952, *Phys. Rev.*, **85**, 216 ; 1954, *Ibid.*, **95**, 321.

PELLAM, J. R., and MORSE, P. M., 1950, *Phys. Rev.*, **78**, 474.

PELLAM, J. R., and SQUIRE, C. F., 1947, *Phys. Rev.*, **72**, 1245.

PESHKOV, V., 1946, *J. Exp. Theor. Phys.*, **16**, 1000 ; *J. Phys., Moscow*, **10**, 389 ; *Proceedings of the Cambridge Low Temperature Conference* (London : Physical Society), p. 19 ; 1952, *J. Exp. Theor. Phys.*, **23**, 687 ; 1954, *Ibid.*, **27**, 351.

Pippard, A. B., 1951, *Phil. Mag.*, **42**, 1209.

Pomeranchuk, I., 1949, *J. Exp. Theor. Phys.*, **19**, 42.

Prigogine, I., 1954, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **3**, 131.

Reekie, J., and Hutchison, T. S., 1953, *Phys. Rev.*, **92**, 827.

Roberts, T. R., and Sydoriak, S. G., 1954, *Phys. Rev.*, **93**, 1418.

Rollin, B. V., 1936, *Thesis*, University of Oxford.

Rollin, B. V., and Simon, F. E., 1939, *Physica*, **6**, 219.

Schiff, L., 1941, *Phys. Rev.*, **57**, 839.

Simon, F. E., 1934, *Nature, Lond.*, **133**, 460, 524 ; 1950, *Physica*, **16**, 753.

Simon, F. E., and Swenson, C. A., 1950, *Nature, Lond.*, **165**, 829.

Smith, B., and Boorse, H. A., 1955, *Phys. Rev.*, **98**, 328.

Sommers, H. S., Dash, J. G., and Goldstein, L., 1955, *Phys. Rev.*, **97**, 855.

Sydoriak, S. G., Grilly, E. R., and Hammel, E. F., 1949, *Phys. Rev.*, **75**, 303.

Taconis, K. W., Beenakker, J. J. M., and Dokoupil, Z., 1950, *Phys. Rev.*, **78**, 171.

Tisza, L., 1938, *C.R. Acad. Sci., Paris*, **207**, 1035, 1186.

Tjerkstra, H. H., 1952, *Physica*, **18**, 853.

Thellung, A., 1953, *Physica*, **19**, 217.

de Troyer, A., van Itterbeek, A., and van den Berg, G. J., 1951, *Physica*, **17**, 50.

Vinen, W. F., 1955, *Conférence de Physique des Basses Températures, Paris* (Paris : Institut International du Froid), p. 60.

de Vries, G., and Daunt, J. G., 1953, *Phys. Rev.*, **92**, 1572 ; 1954, *Ibid.*, **93**, 631.

Ward, J. C., and Wilks, J., 1951, *Phil. Mag.*, **42**, 314 ; 1952, *Ibid.*, **43**, 48.

Waring, R. K., 1955, *Phys. Rev.*, **99**, 1704.

Weinstock, B., Osborne, D. W., and Abraham, B. M., 1949, *Proceedings of the International Conference on Physics of Low Temperatures* (M.I.T.), p. 47.

Wheeler, R. G., Blakewood, C. H., and Lane, C. T., 1955, *Phys. Rev.*, **99**, 1667.

Wilks, J., 1953, *Nuovo Cim.*, Suppl., **10**, 509.

Winkel, P., 1955, *Physica*, **21**, 322.

Winkel, P., Broese van Groenou, A., and Gorter, C. J., 1955, *Physica*, **21**, 345.

Winkel, P., Delsing, A. M. G., and Gorter, C. J., 1955, *Physica*, **21**, 312.

Winkel, P., Delsing, A. M. G., and Poll, J. D., 1955, *Physica*, **21**, 331.

Ziman, J. M., 1953 a, *Proc. Roy. Soc.* A, **219**, 257 ; 1953 b, *Phil. Mag.*, **44**, 548 ; 1954, *Ibid.*, **45**, 100.

APPENDIX

Recent work on liquid ^{4}He not mentioned in earlier reviews and not referred to in the present paper

Melting curve and zero point energy

Hurst, D. G., and Henshaw, D. G., 1955, *Canad. J. Phys.*, **33**, 797.

Swenson, C. A., 1953, *Phys. Rev.*, **89**, 538.

Phonons

van den Berg, G. J., van Itterbeek, A., van Aardenne, G. M. V., and Herfkens, J. H. J., 1955, *Physica*, **21**, 860.

Geilikman, B. T., 1953, *Dokl. Akad. Nauk., S.S.S.R.*, **94**, 191.

van Itterbeek, A., and Forrez, G., 1954, *Physica*, **20**, 133.

Rotons

Beaumont, C. F. A., 1954, *Canad. J. Phys.*, **32**, 759.

Brenig, W., 1956, *Z. Phys.*, **144**, 488.

Goldstein, L., 1953, *Phys. Rev.*, **89**, 597.

Houston, W. V., and Rorscach, H. E., 1955, *Phys. Rev.*, **100**, 1003.

Hurst, D. G., and Henshaw, D. G., 1955, *Phys. Rev.*, **100**, 994.

Kaempffer, F. A., 1954, *Canad. J. Phys.*, **32**, 264.

KLEMENS, P. G., 1955, *Phys. Rev.*, **97**, 1181.
KUPER, C. G., 1955, *Proc. Roy. Soc.* A, **233**, 223.
MATSUDA, H., 1954, *Progr. Theor. Phys., Japan*, **12**, 537.
MIKURA, Z., 1954, *Progr. Theor. Phys., Japan*, **11**, 207.
PEKERIS, C. L., 1953, *Proc. Nat. Acad. Sci., Wash.*, **39**, 443.
RICE, O. K., 1956, *Phys. Rev.*, **102**, 1416.
TEMPERLEY, H. N. V., 1953, *Proc. Phys. Soc.* A, **66**, 995 ; 1954, *Ibid.*, **67**, 901.
TWEET, A. G., 1954, *Phys. Rev.*, **93**, 15.
TYABJI, S. F. B., 1953, *Nature, Lond.*, **172**, 849 ; 1954, *Proc. Camb. Phil. Soc.*, **50**, 449.

Superfluid flow
CHANDRASEKHAR, B. S., and MENDELSSOHN, K., 1953, *Proc. Roy. Soc.* A, **217**, 18.
LANE, C. T., and DYBA, R. V., 1953, *Phys. Rev.*, **92**, 829.

Viscosity
DONNELLY, R. J., CHESTER, G. V., WALMSLEY, R. H., and LANE, C. T., 1956, *Phys. Rev.*, **102**, 3.
GAMTSEMLIDZE, G. A., 1955, *Dokl. Akad. Nauk., S.S.S.R.*, **100**, 441.
HOLLIS-HALLETT, A. C., 1953, *Proc. Camb. Phil. Soc.*, **49**, 717.

Fountain effect
FORSTAT, H., and REYNOLDS, C. A., 1956, *Phys. Rev.*, **101**, 513.
HERIVAL, J. W., 1954, *Nature, Lond.*, **174**, 322.
MANCHESTER, F. D., 1955, *Canad. J. Phys.*, **33**, 146.
VAN DEN MEIJDENBERG, C. J. N., TACONIS, K. W., BEENAKKER, J. J. M., and WANSINK, D. H. N., 1954, *Physica*, **20**, 157.
REYNOLDS, C. A., 1954, *Phys. Rev.*, **93**, 1118.

Heat transport
INOUE, T., 1953, *J. Phys. Soc., Japan*, **8**, 450.
KLEMENS, P. G., 1955, *Aust. J. Phys.*, **8**, 206.
KURTI, N., and MCINTOSH, J., 1955, *Phil. Mag.*, **46**, 104.
MIKURA, Z., 1954, *Progr. Theor. Phys., Japan*, **11**, 503.

Irreversible processes
RICE, O. K., 1953, *Phys. Rev.*, **89**, 793.

Absorption of pressure and temperature waves
CHASE, C. E., 1956, *Amer. J. Phys.*, **24**, 136.
ZINOV'EVA, K. N., 1953, *J. Exp. Theor. Phys.*, **25**, 235.

Mean free path effects
OSBORNE, D. V., 1956, *Phil. Mag.*, **1**, 301.

Supercritical flow
DASH, J. G., 1954, *Phys. Rev.*, **94**, 1091 ; 1955, *Ibid.*, **97**, 1038.
KASUYA, T., 1953, *Progr. Theor. Phys., Japan*, **9**, 87, 89, 90.
SWIM, R. T., and RORSCHACH, H. E., 1955, *Phys. Rev.*, **97**, 25.
WANSINK, D. H. N., TACONIS, K. W., STAAS, F. A., and REUSS, J., 1955, *Physica*, **21**, 596

Rotating helium
ANDRONIKASHVILI, E. L., 1952, *J. Exp. Theor. Phys.*, **22**, 62.
BLATT, J. M., BUTLER, S. T., and SCHAFROTH, M. R., 1955, *Phys. Rev.*, **100**, 481.
BUTLER, S. T., and BLATT, J. M., 1955, *Phys. Rev.*, **100**, 495.
LIFSHITZ, I. M., and KAGANOV, M. I., 1955, *J. Exp. Theor. Phys.*, **29**, 257 ; *Soviet Physics*, **2**, 172.

Film

Franchetti, S., 1955, *Nuovo Cim.*, **2**, 1127.
McCrum, N. G., and Eisenstein, J. C., 1955, *Phys. Rev.*, **99**, 1326.
Rice, O. K., and Widom, B., 1953, *Phys. Rev.*, **90**, 987.

Thickness of the film

Atkins, K. R., 1953, *Phys. Rev.*, **92**, 1571 ; 1954, *Canad. J. Phys.*, **32**, 347.
Bowers, R., 1953, *Phys. Rev.*, **91**, 1016 ; 1953, *Phil. Mag.*, **44**, 1309.
Brewer, D. F., and Mendelssohn, K., 1953, *Phil. Mag.*, **44**, 340.
Jackson, L. C., and Henshaw, D. G., 1953, *Phil. Mag.*, **44**, 14.
Knudsen, W. C., and Dillinger, J. R., 1954, *Phys. Rev.*, **95**, 279.
McCrum, N. G., 1954, *Phil. Mag.*, **45**, 1302.

Transport properties of the film

Beams, W. J., 1955, *Phys. Rev.*, **98**, 1138.
Dyba, R. V., Lane, C. T., and Blakewood, C. H., 1954, *Phys. Rev.*, **95**, 1365.
Esel'son, B. N., and Lazarev, B. G., 1952, *J. Exp. Theor. Phys.*, **23**, 552.
Ham, A. C., and Jackson, L. C., 1953, *Phil. Mag.*, **44**, 1403.
Picus, G. S., 1953, *Phys. Rev.*, **90**, 719 ; 1954, *Ibid.*, **94**, 1459.
Taconis, K. W., 1955, *Phys. Rev.*, **97**, 1176.

The unsaturated film

Bowers, R., 1953, *Phil. Mag.*, **44**, 485.
McCrum, N. G., and Mendelssohn, K., 1954, *Phil. Mag.*, **45**, 102.
Tjerkstra, H. H., Hooftman, F. J., and van den Meydenberg, C. J. N., 1953, *Physica*, **19**, 935.
White, D., Chou, C., and Johnston, H. L., 1952, *J. Chem. Phys.*, **20**, 1819.

^{3}*He*–^{4}*He solutions*

Guttman, L., and Arnold, J. R., 1953, *Phys. Rev.*, **92**, 547.
Khalatnikov, I. M., 1952, *J. Exp. Theor. Phys.*, **23**, 264.

The nature of the lambda transition

Blatt, J. M., and Butler, S. T., 1954, *Phys. Rev.*, **96**, 1149 ; 1955, *Ibid.*, **100**, 476.
Buckthought, K., 1953, *Canad. J. Phys.*, **31**, 932.
Chester, G. V., 1954, *Phys. Rev.*, **93**, 1412, **94**, 246.
Franchetti, S., 1955, *Nuovo Cim.*, **12**, 743, *Ibid.*, Suppl. **1**, 159.
Goldstein, L., 1955, *Phys. Rev.*, **100**, 981.
Goldstein, L., and Sommers, H. S., 1956, *Phys. Rev.*, **101**, 1235.
Hurst, C., 1955, *Proc. Phys. Soc.* B, **68**, 521.
Kikuchi, R., 1954, *Phys. Rev.*, **96**, 563.
Lawson, A. W., and Meyer, L., 1954, *Phys. Rev.*, **93**, 259.
van der Leeden, P., 1955, *Physica*, **21**, 561.
Price, P. J., 1954, *Phys. Rev.*, **94**, 257.
Rice, O. K., 1954, *Phys. Rev.*, **93**, 1161, **96**, 1460.
Zilsel, P. R., 1953, *Phys. Rev.*, **92**, 1106.

Miscellaneous

Atkins, K. R., 1953, *Canad. J. Phys.*, **31**, 1165.
Berman, R., and Poulter, J., 1952, *Phil. Mag.*, **43**, 1047.
Clement, J. R., Logan, J. K., and Gaffney, J., 1955, *Phys. Rev.*, **100**, 743.

HIGH-ENERGY PROTONS

By A. E. TAYLOR

Atomic Energy Research Establishment, Harwell, Didcot, Berks.

CONTENTS

PAGE
- § 1. Introduction 87
- § 2. Proton–proton scattering 88
  - 2.1. Introduction 88
  - 2.2. At energies below 400 MeV 88
    - (a) The differential cross section at 90° 88
    - (b) Nuclear–coulomb interference 90
    - (c) Polarization 91
      - (i) Double scattering 91
      - (ii) Sign of the polarization 95
      - (iii) Triple scattering 96
    - (d) Phase shift analysis 96
      - (i) Proton–proton scattering 97
      - (ii) Neutron–proton scattering 99
      - (iii) Comparison with models 101
    - (e) Scattering matrix 103
    - (f) Charge symmetry and independence 103
    - (g) Conclusion 104
  - 2.3. At energies above 400 MeV 104
    - (a) Total cross section 104
    - (b) Elastic scattering 105
    - (c) Inelastic scattering and meson production 106
    - (d) Antiprotons 108
- § 3. Proton–nuclear scattering 110
  - 3.1. Introduction 110
  - 3.2. Nuclear transparency and the Goldberger model 110
  - 3.3. Inelastic scattering groups 112
  - 3.4. Internal momentum distribution 114
  - 3.5. Optical model 115
    - (a) Introduction 115
    - (b) Total and differential cross sections 116
    - (c) Absorption cross sections and nuclear radii 118
  - 3.6. Polarization 119
    - (a) Introduction 119
    - (b) Elastic scattering and spin–orbit coupling 120
    - (c) Variation with energy 120
    - (d) Variation with scattering angle 122
    - (e) Variation with atomic weight 123
    - (f) Coulomb interference 124
    - (g) Imaginary part of spin–orbit potential 125
- Acknowledgments 126
- References 126

Abstract. Recent experiments with high-energy protons above 100 MeV are reviewed. Considerable data have been obtained from the study of proton–proton scattering below 400 MeV, especially where polarized beams have been used. It is now possible to extract the available information from such experiments by making phase shift analyses. The neutron–proton scattering experiments, although not so numerous, are reviewed where they are relevant to these analyses. The phase shifts have not yet been determined unambiguously but the analyses have indicated the general trend and make it possible to test various models of the nucleon–nucleon interaction. At energies above 400 MeV the character of the scattering changes and the production of mesons becomes noticeable. At the highest energies now available from the Bevatron at Berkeley antiprotons are produced.

Many of the experimental results in proton–nuclear scattering have been explained in terms of nuclear transparency, optical and statistical models, but careful measurements of inelastic scattering have suggested that more detailed models are required. The polarization in elastic scattering can be explained by an optical model with the addition of a spin–orbit coupling similar to that used in the shell model. The theoretical variation of the polarization with scattering angle displays sharp features near the diffraction minima but the observation of these tends to be obliterated by finite experimental angular resolution and detection of inelastically scattered particles.

---

§ 1. INTRODUCTION

A MORE complete knowledge of the nucleon–nucleon interaction is of prime importance in the further understanding of nuclear physics. One of the most direct ways of gaining this knowledge is to investigate nucleon–nucleon scattering. The final programme would be to calculate the cross sections from field theory, but as this is beset with many difficulties, phenomenological descriptions have been proposed. For instance the interaction has been represented by a static potential, which may be spin and charge dependent. For any particular model it is possible to calculate the phase shifts of the partial waves, and by comparison with those experimentally determined to test the model. The highest states with orbital angular momentum L which are involved in the interaction are those for which $L \sim kR$ where k is the relative wave number and R the range of the interaction. For triplet states there are three phase shifts and one coupling parameter for $L > 1$, and for singlet states one phase shift. The unpolarized cross section for proton–proton scattering can be described by $L+1$ parameters, i.e. the coefficients of the Legendre polynomial. Hence it is only possible to determine from this cross section the phase shift when $L = 0$, i.e. at low energies. It is only recently that it has been possible to attempt to determine the other phase shifts experimentally. Unpolarized beams are a mixture of the various spin states and information is lost through averaging over the initial and final spin states. However, the production of polarized beams has allowed additional information to be obtained. The cross sections will be different from those for an unpolarized beam if the interaction couples spin and spatial coordinates. Much of the theoretical formalism for such experiments has been developed and recently reviewed by Wolfenstein (1956).

Experiments at high energies are also particularly useful as a probe of the internal nuclear structure since the wavelength of the incident particle is then less than the inter-nucleon spacing and the nuclei have been found to be partially transparent. Many of the gross features of the results have been interpreted in terms of this transparency and of an optical analogy. The more detailed results on inelastic scattering have required consideration of the individual nuclear states. As a consequence of the spin dependency of the nucleon–nucleon interaction and of the transparency of the nucleus, polarization effects are also expected in the nucleon–nuclear scattering. These effects have been explained also by invoking a similar spin–orbit coupling to that used in the shell model of the nucleus.

Most of the detailed measurements have been made with protons since they are experimentally easier to use. For this reason the review will be mainly concerned with proton scattering. Nevertheless it will be found convenient to include some of the results of the neutron scattering experiments.

§ 2. PROTON–PROTON SCATTERING

2.1. *Introduction*

It had been hoped that since the high energies available from accelerating machines bring in higher angular momentum states in the scattering, it would be possible to give a more reasonable and detailed description of the p–p interaction than had been obtainable in the region of 10 MEV. Here only s states contribute to the scattering and no information can be obtained about the shape of potential. In the last review on nucleon–nucleon scattering by Christian in *Reports on Progress in Physics* (1952) the results then available at energies up to 300 MEV were analysed in terms of static potentials. Since then higher energy machines have come into operation extending the field to 6000 MEV. Also more detailed work has been done in the region of 100–400 MEV where, as shown by Christian, the phenomenological approach of static potentials gave only fair agreement with the experimental results, and where with this approach it is difficult to maintain the hypothesis of charge independence of nuclear forces. Agreement has now been achieved between the various laboratories on the absolute value of the differential cross section for p–p scattering at 90° c.m. in the region from 100 to 300 MEV. In this energy range the scattering is nearly isotropic, but at higher energies this is no longer the situation and inelastic scattering takes place with the production of mesons. One of the most notable advances in recent years has been the production of polarized beams. This has enabled studies to be made of the non-central character of the nucleon–nucleon interaction in the triplet states. The emphasis in the theoretical interpretation of all these results has now shifted from a phenomenological approach using static potentials to one of phase shift analysis, with the hope that when all the phase shifts are known as a function of energy it will then be possible to construct the appropriate potentials or to test any particular model by the phase shifts it predicts. Although meson production is energetically possible at 290 MEV the character of the p–p scattering changes noticeably only above 400 MEV. Hence it will be found convenient to review first the experiments below 400 MEV.

2.2. *Proton–Proton Scattering at Energies below* 400 MEV

(*a*) *The differential cross section at* 90°.

Early measurements of the differential cross section for p–p scattering had shown that near-isotropy exists in the scattering. However, recent measurements at Harvard at 95 MEV (Kruse, Teem and Ramsey 1956) have shown a significant decrease from a value of 4·9 mbn sterad^{-1} at 40° to 4·6 mbn sterad^{-1} at 90°. (Energies refer to incident particle energy in the laboratory system, and angles and cross sections to the centre-of-mass system.) Also the angular distributions previously reported between 100 and 260 MEV are quite consistent with a similar variation which can be put in the form

$$\left(\frac{\partial\sigma}{\partial\Omega}\right)_\theta = \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}(1 + a\cos^2\theta). \qquad \ldots\ldots(1)$$

Analysis outside the region of the nuclear–coulomb interference of all these results, and the later work of Garrison (1954) and of Chamberlain and Garrison (1954) at 170 and 260 MEV, leads to a value of *a* lying between 0·1 and 0·2, as can be seen from table 1.

Table 1. Differential p–p Cross Section between 95 and 260 MEV

Energy (MEV)	Differential cross section (mbn sterad^{-1})	Reference
95	$4{\cdot}60\,(1+(0{\cdot}12\pm0{\cdot}04)\cos^2\theta)$	1
147	$3{\cdot}64\,(1+(0{\cdot}08\pm0{\cdot}07)\cos^2\theta)$	2
170	$3{\cdot}14\,(1+(0{\cdot}19\pm0{\cdot}04)\cos^2\theta)$	3
170	$3{\cdot}47\,(1+(0{\cdot}18\pm0{\cdot}03)\cos^2\theta)$	3
260	$3{\cdot}50\,(1+(0{\cdot}11\pm0{\cdot}02)\cos^2\theta)$	3

1, Kruse, Teem and Ramsey 1956 ; 2, Cassels, Pickavance and Stafford 1952, renormalized ; 3, Chamberlain and Garrison 1954, Garrison 1954.

There has been until recently a discrepancy between the values of $(\partial\sigma/\partial\Omega)_{90}$ obtained at Rochester and Harwell and those obtained at Chicago and Berkeley. This has now been largely resolved, in part due to redeterminations and in part due to a re-measurement of the ^{12}C (p, pn)^{11}C cross section (Birnbaum, Crandall, Millburn and Pyle 1954) which had been used in some of the measurements to monitor the incident number of protons. The previously measured values of this cross section have been shown to be in error by as much as 20%. The Rochester values of $(\partial\sigma/\partial\Omega)_{90}$, although not fully dependent on the above measurements, do need to be revised and are now $4{\cdot}05\pm0{\cdot}35$ mbn sterad^{-1} (Oxley and Schamberger 1952) and $3{\cdot}80\pm0{\cdot}32$ mbn sterad^{-1} (Towler 1951). Cassels (1956) has also re-measured the value at 147 MEV obtaining the value $4{\cdot}05\pm0{\cdot}20$ mbn sterad^{-1}. In the previous measurement (Cassels, Pickavance and Stafford 1952) of 4·90 mbn sterad^{-1} an independent method of monitoring the flux was used as well as the ^{12}C (p, pn) ^{11}C cross section. As agreement was obtained between the two methods, presumably some untraced error was present in the independent method. Recent measurements at Harvard (Kruse, Teem and Ramsey 1956) have been made in the range 41 to 95 MEV, and have shown the 90° differential cross section to fall smoothly from a value of 11 mbn sterad^{-1} at 41 MEV to 4·60 mbn sterad^{-1} at 95 MEV.

The above measurements were a direct determination of the 90° differential cross section. Several measurements have been made of the total p–p cross section and a value of the 90° differential cross section inferred. It was thought at the time that one of the causes of the discrepancy in the direct measurements might have arisen from the calibration of the ionization chamber or the use of the ^{12}C (p, pn) ^{11}C cross section. The measurements of the total cross section provide an independent check on the 90° differential cross section as they do not rely upon these methods of measuring the incident flux but count the number of incident protons directly.

The determinations of the total cross section have involved absorption measurements in liquid hydrogen, and are strictly the determination of the integrated differential cross section for scattering through angles greater than some minimum angle $\theta_{\min}$. If isotropy of the scattering is assumed then this integrated cross section $\sigma(\theta_{\min})$ is related to the 90° differential cross section $(\partial\sigma/\partial\Omega)_{90}$ by the relation

$$\sigma(\theta_{\min}) = 2\pi \cos \theta_{\min}(\partial\sigma/\partial\Omega)_{90}. \qquad \ldots\ldots(2)$$

It is this isotropic or average differential cross section which Chamberlain, Pettengill, Segrè and Wiegand (1954) have reported at 330 and 225 Mev. If, however, the angular distribution is that given by equation (1), this average value at 225 Mev should be decreased slightly to obtain the 90° value. The amount of the decrease can be judged by the results at 134 Mev (Taylor and Wood 1956 a), also obtained by an absorption method in liquid hydrogen. Here the value of $(\partial\sigma/\partial\Omega)_{90}$ of $3{\cdot}79 \pm 0{\cdot}13$ mbn sterad^{-1} was obtained assuming isotropy whereas the value is $3{\cdot}64 \pm 0{\cdot}15$ mbn sterad^{-1} with the variation suggested by equation (1). A similar correction has been applied to the 225 Mev result of Chamberlain, Pettengill *et al.* (1954).

The values obtained by both direct and indirect methods are in very good agreement with each other as is seen in figure 1. It will be noted that there is a remarkable constancy of $(\partial\sigma/\partial\Omega)_{90}$ between 130 and 400 Mev.

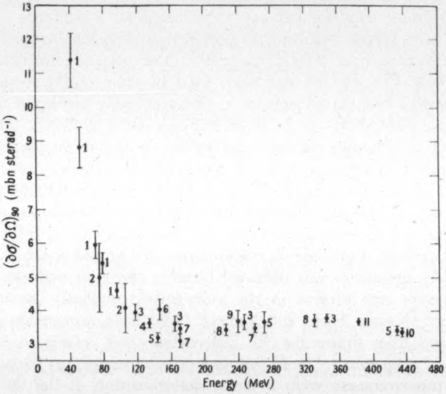


Figure 1. Variation with energy of the differential p–p cross section at 90°. References to points : (1) Kruse, Teem and Ramsey 1956 ; (2) Birge, Kruse and Ramsey 1951, corrected for change in ^{12}C (p, pn) ^{11}C cross section ; (3) Chamberlain, Segrè and Wiegand 1951 ; (4) Taylor and Wood 1956 a ; (5) Marshall, Marshall and Nedzel 1953 ; (6) Cassels 1956 ; (7) Garrison 1954 ; (8) Chamberlain, Pettengill, Segrè and Wiegand 1954 ; (9) Oxley and Schamberger 1952, corrected for change in ^{12}C (p, pn) ^{11}C cross section ; (10) Mott, Sutton, Fox and Kane 1953 ; (11) Holt 1956.

(*b*) *Nuclear–coulomb interference.*

Measurements were usually restricted until recently to scattering angles greater than 30° but with the advent of large supplies of liquid hydrogen in the nuclear physics laboratories, measurements have now been extended to much lower angles. The previous restriction arose because a polyethylene–carbon difference method was used to obtain the hydrogen scattering and for small scattering angles the contribution to the counting rate from the hydrogen was only a small fraction of that from the carbon. The background when liquid hydrogen is used arises from the walls of the container which can be either very thin metal or expanded plastic. However, even with these targets there is a restriction due to the large coulomb scattering from the walls and for measurements at very small angles and low energies the detector has to be arranged so that it does not see those parts of the walls which are irradiated with the incident beam. Measurements can then be made in principle as low as the multiple coulomb scattering and angular divergence of the incident beam will allow.

Typical sets of data are shown in figures 2 and 3 for scattering at 330 Mev (Fischer and Goldhaber 1954) and at 142 Mev (Taylor and Wood 1956 b). The solid curve is the sum of nuclear and coulomb scattering and it is at once apparent that there is a large destructive interference which extends to 20° in contrast with the data of Towler (1951) at 240 Mev and of Chamberlain and Garrison (1954) at 170 and 260 Mev which show a much smaller destructive interference. How far this interference appears to extend depends to some extent on the assumptions that are made about the purely nuclear scattering angular distribution and the relativistic corrections for the coulomb scattering. If the former is that of equation (1), then the interference may extend further than figure 2 suggests, and indeed the value of *a* in equation (1) for purely nuclear scattering may be larger than the angular distribution measurements suggest. This has a bearing on the theoretical interpretation which will be mentioned later.

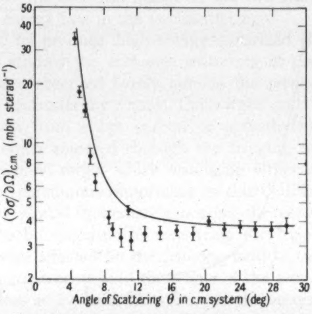


Figure 2. Small angle p–p cross section at 330 Mev. The experimental points are those of Fischer and Goldhaber (1954). The curve is the sum of the coulomb cross section and a constant nuclear cross section of $3\cdot70$ mbn sterad^{-1}.

(*c*) *Polarization.*

(i) *Double scattering.* Perhaps the most interesting feature of p–p scattering to be investigated recently has been the polarization produced in a single scattering and detected by the azimuthal asymmetry in a second scattering. The presence of non-central forces at low energies is shown by the existence of the deuteron quadrupole moment which requires a tensor force in the interaction and by the spin–orbit interaction which has been invoked in the shell model of the nucleus (Mayer 1950). Also the resonance scattering of protons from helium at energies of a few Mev has been shown (Heusinkveld and Freier 1952) to produce polarization. Mott (1932) had suggested a double scattering experiment for electrons. At high energies the various phenomenological models of Jastrow (1951), Case and Pais (1950) and Christian and Noyes (1950), which had been proposed in an attempt to explain the remarkable near isotropy in the p–p scattering, all involved non-central force terms. As a consequence polarization effects were to be expected in nucleon–nucleon scattering at high energies. The degree to which nuclear scattering polarizes the particles, or aligns their spins, will be governed by the spin dependence of nuclear forces. Considerable effort has therefore been expended in the experimental and theoretical study of polarization in the hope of further understanding the nucleon–nucleon interaction. The general formalism for such experiments has been given by Wolfenstein (1949).

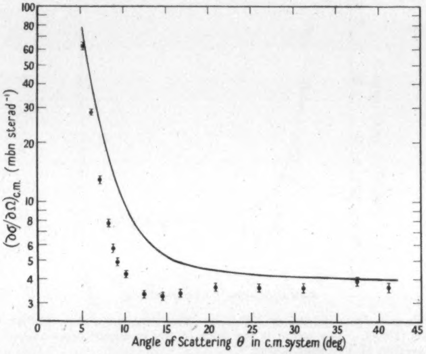


Figure 3. Small angle p–p cross section at 142 Mev. The experimental points are those of Taylor and Wood (1956 b). The curve is the sum of the coulomb cross section and a nuclear cross section = $3{\cdot}70\,(1+0{\cdot}14\cos^2\theta)$ mbn sterad^{-1}.

For particles of spin $\frac{1}{2}$ the polarization $\mathbf{P}$ is a vector of magnitude twice the expectation value of the spin, and therefore varies from 0 to 1. The component of $\mathbf{P}$ in a given direction is equal to $(N_1-N_2)/(N_1+N_2)$, where N_1 and N_2 are the numbers of particles orientated in or opposite to the chosen direction. This direction is perpendicular to the plane of scattering. The analyser used to detect the polarization $\mathbf{P}_1$ produced in the first scattering is usually a second scatterer which can also polarize the beam and which for an unpolarized incident beam would give a polarization $\mathbf{P}_2$. It can then be shown that the intensity in the second scattering is given by

$$I(\theta) = I_0(\theta)[1+P_1P_2\cos\phi]$$

where θ is the scattering angle at the second target, ϕ is the azimuthal angle between the scattering planes and $I_0(\theta)$ is the intensity for an unpolarized beam. Hence by measuring the intensities at the two azimuthal angles $\phi = 0$ and 180°, the asymmetry as usually defined is given by

$$\Sigma = \frac{I(\phi = 0) - I(\phi = 180^\circ)}{I(\phi = 0) + I(\phi = 180^\circ)} = P_1 P_2.$$

Under certain circumstances it is possible to assume that the two scatterings are identical, and then $P = P_1 = P_2 = \pm \Sigma^{1/2}$. A separate experiment, however, has to be made in order to determine the sign of P.

The polarized beams used in studying nucleon–nucleon polarization have usually been obtained by scattering the internal proton beam of the cyclotron by carbon and beryllium targets at angles from a few degrees to 30°. The polarization produced in proton–nuclear scattering is dealt with in § 3.6. There it is shown that great care is needed to determine accurately the value of P_1 as inelastic scattering does not produce the same polarization as does elastic scattering and there may be some uncertainty about the identical nature of the two scatterings because of this, quite apart from the energy loss in the two scatterings.

An early attempt to produce high-energy polarized nucleons was made by Wouters (1951), who studied the exchange scattering of protons in LiH and LiD. Only small effects were observed barely outside the errors. The first successful attempt was that at Rochester by Oxley, Cartwright and Rouvina (1954). Here the first scattering was from either carbon or polyethylene inside the cyclotron and the polarized protons emerged through the fringing field of the cyclotron to be scattered from a second target which was again either carbon or polyethylene. Since intensity was of paramount importance in this double-scattering experiment, the second target was placed immediately outside the vacuum tank. This meant that great care had to be exercised in measuring the trajectories of the scattered particles since they were affected by the fringing field of the cyclotron. Magnetic shielding was therefore placed around the whole of the second scattering apparatus. The first scattering was at 19° and 225 MEV, and the second scattering was at 27° and 200 MEV. Measurements were made for the four azimuthal angles 0°, 90°, 180° and 270°. The variation of polarization with angle and energy was neglected over these ranges and the polarization produced in p–p scattering was estimated to be $22 \pm 4\%$.

Since this first demonstration of the effect at Rochester in 1953, many laboratories with machines in the 100 to 400 MEV energy range have undertaken studies of polarization. The experimental details may vary slightly but the main pattern remains the same. Carbon and beryllium have been used almost exclusively as the first target to produce the polarized beam, which has either been allowed to spill out of the cyclotron or has been guided out by magnetic channels—the angle of the first scattering being within the range 6° to 20°. The beam may then be guided by bending magnets and focused to a position well outside the fringing field of the cyclotron. Fluxes have been obtained of 10^4 to 10^6 protons $\text{cm}^{-2}\,\text{sec}^{-1}$ with polarizations varying from 50 to 90%. For the second scatterer, polyethylene or liquid hydrogen has been used. Measurements at the two azimuthal angles $\phi = 0$ and 180° have been made either by rotating the counting apparatus around the direction of the polarized beam or by moving an arm on which the apparatus is mounted from one side of the beam to the other. In both methods extreme care is needed in aligning the apparatus as slight misalignments can produce false asymmetries of the same order as the measured effect. For instance, in a typical experiment it might be found necessary to align the apparatus to 1 minute of arc if the false asymmetries are to be less than 1% (the measured asymmetries may be only a few per cent). A method which overcomes some of these difficulties has been used by Hillman, Stafford and Whitehead (1956). Here the spin direction was rotated by sending the beam through a longitudinal magnetic field which causes the magnetic moment to precess. Although this was used for neutrons it could also be used for protons as there is no net deflection of the protons. Further, when using a polyethylene–carbon subtraction method to obtain the p–p polarization, it is necessary to ensure that there is the same amount of inelastic scattering detected from the carbon content of the polyethylene as from the carbon scatterer.

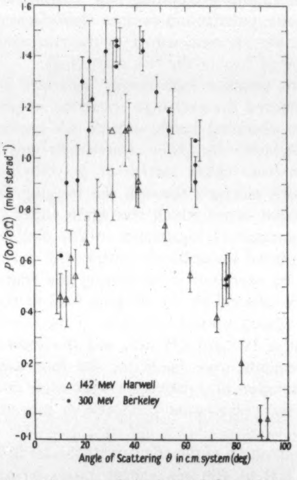


Figure 4. p–p polarization at 142 and 300 Mev.

This is particularly difficult when the scattering angle is just a little less than the first diffraction minimum. At this angle the elastic scattering polarization is large but decreasing rapidly as the diffraction minimum is approached where the relative intensities of elastic and inelastic scattering are comparable (see § 3.3). This difficulty arises again in the determination of the incident beam polarization by a double-scattering from carbon since the internal beam of a cyclotron is not mono-energetic. One method of overcoming this difficulty to some extent, at the expense of some degree of polarization, is to use scattering angles in the coulomb–nuclear interference region where both cross section and polarization are largely independent of angle (see figures 15 and 20) and where the inelastic events are a small portion of the total scattering. Polarizations of the order of 50% can be achieved by this method in contrast to the higher values of 80 to 90% obtained by scattering at larger angles.

Measurements have been made on the p–p polarization in the energy range 100 to 400 MEV (Dickson and Salter (1954) and Taylor and Wood (1956 b), at Harwell; Oxley, Cartwright and Rouvina (1954), Baskir, Tinlot and Spital (1955) and Baskir (1956, private communication), at the University of Rochester; de Carvalho, Heiberg, Marshall and Marshall (1954) and Marshall, Marshall and de Carvalho (1954), at the University of Chicago; Chamberlain, Segrè, Tripp, Wiegand and Ypsilantis (1954), Pettengill (1954), Ypsilantis (1955) and Fischer and Baldwin (1955), at Berkeley; Kane, Stallwood, Sutton, Fields and Fox (1954), at the Carnegie Institute of Technology). A typical set of experimental results is shown in figure 4. Most of these measurements have been analysed in the form (Wolfenstein 1949)

$$P\left(\frac{\partial\sigma}{\partial\Omega}\right) = \sin 2\theta[a+b\cos^2\theta+c\cos^4\theta+\ldots] \qquad \ldots\ldots(3)$$

where P is the measured polarization and $\partial\sigma/\partial\Omega$ the unpolarized cross section. Some of the measurements at Berkeley (Chamberlain, Donaldson, Segrè, Tripp, Wiegand and Ypsilantis 1954) have utilized the quasi p–p scattering in p–d scattering to obtain the polarization. The values of the constants in equation (3) are listed in table 2.

Table 2. Angular Variation of Polarized p–p Cross Section

Energy† (MEV)	Laboratory	Constants of equation (3) (mbn sterad^{-1}) *a*	*b*	*c*
132	Rochester	0·42 ± 0·11	0·42 ± 0·18	
142	Harwell	0·555 ± 0·057	0·64 ± 0·10	
170	Rochester	0·78 ± 0·10	0·30 ± 0·18	
170	Berkeley	0·56 ± 0·16	0·54 ± 0·25	
210	Rochester	1·03 ± 0·08	0·32 ± 0·13	
300	Berkeley	1·02	0·82	
315	Berkeley	0·99	0·74	
415	*Carnegie	1·11	1·1	0·70

* The Carnegie results were normalized to $(\partial\sigma/\partial\Omega)_{90} = 1{\cdot}0$ and in the table have been multiplied by the value 3·50 mbn sterad^{-1}.

† Energy of incident proton beam.

(ii) *Sign of the polarization.* As was mentioned earlier, the above double-scattering experiments give no information about the sign of the polarization. In order to determine this it is necessary to use some process which is either known or calculable. Both at Chicago (Marshall and Marshall 1955) and at Harwell (Brinkworth and Rose 1956) the scattering from He was chosen as the analyser for a double-scattering experiment. The beams were slowed down to a few MEV and the resultant asymmetry in the scattering was measured. Different geometries were used, but the experiments were essentially the same. There is strong evidence to believe that depolarization does not occur in the slowing down process (Heiberg, Kruse, Marshall, Marshall and Solmitz 1955, Wolfenstein 1949). The angular distribution of protons scattered from helium in the range 1 to 10 MEV has been measured and analysis of the results provided a unique set of phase shifts (Heusinkveld and Freier 1952). Hence it was possible to calculate the polarization in proton–helium elastic scattering and from the measured asymmetry to determine the sign of the polarization produced in the first scattering. In the convention normally adopted this measurement gave the polarization as positive, i.e. spin-up particles scatter predominantly to the left. The positive direction is defined as that of $\mathbf{k} \times \mathbf{k}'$ where $\mathbf{k}$ and $\mathbf{k}'$ are the propagation vectors of the incident and scattered waves respectively.

(iii) *Triple scattering.* Wolfenstein (1954) has pointed out that further information on the interaction can be obtained from triple-scattering experiments. In these experiments the change in polarization caused by the scattering is measured. The first scattering serves to polarize the beam, the second scattering can change the direction and/or magnitude of the polarization and the third scattering serves to analyse the beam from the second scattering. A simple analyser only determines the final component of $\mathbf{P}$ perpendicular to the scattering plane. Therefore if all the scatterings occur in the same plane, the asymmetry in the third scattering is given (Wolfenstein) by the relation

$$\Sigma = \frac{P_3(P_2 + DP_1 \cos\phi)}{1 + P_1 P_2 \cos\phi}$$

where P_i refers to the polarization in the ith scattering. If the various P_i's are known, then a measurement for either $\phi = 0$ or $180°$ is sufficient to determine D. For a completely polarized incident beam in which $P_1 = 1$, then D must equal unity if no depolarization occurs. Hence D is a measure of the probability of the polarization remaining unchanged. If the scatterings occur in planes successively perpendicular, then the asymmetry in the third scattering is given by the relation

$$\Sigma = P_3 P_1 R \sin\phi$$

where $\phi = 90°$ or $270°$ and $P_1 R$ is the component perpendicular to the final plane of scattering. More complicated experiments are possible using an auxiliary magnetic field such that the direction of the field is perpendicular both to the proton spin and its velocity. Since the proton has an anomalous magnetic moment, the spin is rotated through a greater angle than the beam. Hence there is a component of spin along the direction of the beam, and measurements of asymmetry enable another independent parameter A to be obtained (Wolfenstein 1954). Several of these types of triple-scattering measurements have been made at Berkeley for 310 Mev protons (Ypsilantis, Wiegand, Tripp, Segrè and Chamberlain 1955) and a typical set of results for the depolarization parameter D is shown in figure 5.

Yet another experiment which would yield more information on the interaction has been suggested by Wolfenstein (1956), Stapp (1955) and Oehme (1955). This would involve the analysis in coincidence of the relative polarization states of the scattered and recoil protons. As yet this experiment has not been made.

(*d*) *Phase shift analysis.*

Of the various static potential models, only that of Jastrow (1951) maintained the near isotropy in p–p scattering to sufficiently small angles. However, the later analysis of Goldfarb and Feldman (1952) and Swanson (1953) showed that only the *L–S* model of Case and Pais (1950) gave the right magnitude of the observed polarizations. Unfortunately the same analysis showed that a more exact treatment of this model destroyed the qualitative agreement with the p–p differential cross section. In view of the apparent failure of this phenomenological approach, the available information contained in the experimental data is now almost exclusively extracted by way of a phase shift analysis. Any particular model can then be checked by the agreement it gives with the phase shifts so determined. Further, when all the phase shifts are known as a function of energy there is the hope that an appropriate model can be constructed.

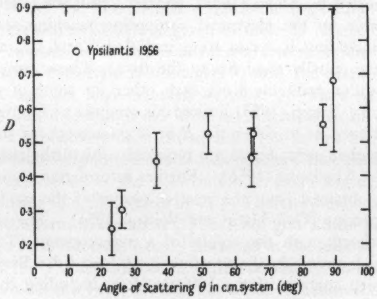


Figure 5. Depolarization parameter at 310 Mev.

(i) *Proton–proton scattering*. The number of independent pieces of information which can be obtained from the differential cross section is $L+1$ where L is the order of the highest angular momentum state participating in the scattering. For each triplet state with $L>1$ the number of associated parameters is four: three phase shifts and one mixture or coupling parameter. Hence it will be seen that when the higher angular momentum states are participating, many more pieces of independent information are required than those provided by the differential cross section. These come from the polarization experiments described above from the angular dependency of P and the D, R and A parameters. Additional information also comes from the coulomb–nuclear interference region. For instance, in the work at Berkeley at 300 Mev a restriction might be made of $L \leqslant 3$ (Stapp 1955). Then eight phase shifts and one coupling parameter are possible and only four pieces of information are contained in the differential cross section; three more can come from the polarization data in the double-scattering experiment. In order to make a solution possible, further information is required and this is obtained via the triple-scattering and magnetic-field experiments. Of course, further information still is obtained from the coulomb–nuclear interference, but as this is particularly sensitive to the terms due to higher L values, it has been used mainly as a check on solutions obtained without its inclusion. The sensitivity arises since the interference is a cross product of the nuclear and coulomb scattering amplitudes. At lower energies, say 100 Mev, the triple-scattering experiments become very difficult as the energy of the scattered beam is too low for any analyser to give measurable asymmetries. The interference region can provide the additional information from both the cross section and the polarization measurements. The difficulty which arises is of knowing the exact relativistic form for the coulomb terms and the part played by the anomalous magnetic moment (Garren 1955). If the experiments are exact enough, and this applies particularly to the measurement of the polarization at small angles, then this interference region can be used to find the extent of the contribution of high-order terms and so provide a unique set of phase shifts at the lower energies.

The most complete analysis of the Berkeley 300 MEV data along these and similar lines has been reported by Stapp (1955), and by Ypsilantis (1956), who processed all the data available on the electronic computing machine MANIAC. All phase shifts up to and including h waves were included, i.e. $L \leqslant 5$, and two sets were obtained which gave equally good fits to the data. These sets, although roughly in agreement, do differ markedly from each other for some of the partial waves, as is seen in table 3. Stapp (1955) limited his analysis to f waves and found that he could not get a good fit to either the R or A parameter for small angles, which suggests that the higher order terms are required. Nevertheless, his best solution is similar to those of Ypsilantis (1956). Further information about the phase shifts at 300 MEV can be obtained from the relative phases of the modes of meson production in p–p scattering (Gell-Mann and Watson 1954). The solutions found by Ypsilantis are consistent with the results of a measurement (Tripp 1955) of the polarization of the deuteron in the reaction $p+p \rightarrow \pi^+ + d$. Breit (1956) has also reported an extensive analysis of the 300 MEV data, including that from Chicago, again having resort to an electronic computer. Similar analyses have been made of lower energy data by Thaler, Bengston and Breit (1954) at 240 MEV, by Ohnuma and Feldman (1956) at 150 MEV, by Garren (1956) at 240 MEV and by Phillips (1956, private communication) at 95 MEV. These analyses have not been so complete as at 300 MEV because there is not such a great wealth of data at the lower energies. However, the analyses have been restricted to lower order waves. It should be pointed out that if the interference region is sensitive to higher L values, then the neglect of these terms in this region may lead to an incorrect solution. Further, the interference is relatively greater and extends to larger angles at lower energies.

In the work of Thaler, Bengston and Breit (1954) the analysis was restricted to the unpolarized cross sections but n–p scattering was also considered and the charge independence hypothesis was invoked. Only terms up to $J=2$ were considered but with a coupling of the $^3P_2$ to the $^3F_2$ scattering. This coupling brings in the effect of states of higher L into the low-energy scattering and alters the phase shifts considerably from those which would be obtained without its inclusion. The value of the $^1S_0$ phase shift was calculated for an assumed meson well. Reasonable agreement with the experimental angular distribution in p–p and n–p scattering was obtained. The analysis of Ohnuma and Feldman of the 150 MEV p–p data was also restricted to $J \leqslant 2$ but no coupling was taken into account. Now the experimental results which have been analysed at 140 MEV in the form of equation (3) have clearly shown (table 2) that there is a sizable term in $b \cos^2 \theta$ which can arise from the normal ^3P–^3F interference even without postulating an additional coupling parameter. Ohnuma and Feldman have therefore chosen the $^3F_2$ phase as non-zero in order to produce this interference. As in the analysis of the 300 MEV results, the polarization data and scattering for $\theta > 30°$ are used to determine sets of the six phase shifts, and the nuclear–coulomb interference unpolarized cross section data are used to make a selection from these. The treatment in this interference region, as pointed out by the authors, is non-relativistic and indeed the nuclear–coulomb interference has been ignored completely in analysing the polarization data. While this may be justified for polarization effects at angles greater than 30°, the effects may be large below this angle. Garren developed the analysis for s and p waves and considered the anomalous magnetic moment as well as the relativistic form of the coulomb interaction. The scattering data at 240 Mev were analysed and one set of the phase shifts which is in approximate accord with both the p–p and n–p experimental data is given in table 3. Since only s and p waves were used it was impossible to obtain any $b \cos^2 \theta$ term in the polarization (equation (3)). If Garren's analysis is used at 140 Mev, the phase shifts obtained are somewhat similar and are also included in table 3. Here again the solution is unreal as there is still a sizable $b \cos^2 \theta$ term. Attempts to fit the results without this term lead to a probability on a chi-squared test as low as 0·5%.

However, it should be pointed out at this juncture that with these phase shifts at 140 Mev the interference in the polarization at small angles is appreciable and destructive, amounting to a 25% effect at 15°. Therefore the analysis of the experimental results in the form of equation (3) does not give a true picture of the purely nuclear terms. In fact, the term $b \cos^2 \theta$ obtained in the least squares analysis is an under-estimate of ^3P–^3F interference. For this reason the solution F selected from those given by Ohnuma and Feldman was that which reproduced the polarization data at large angles but gave an over-estimate at the smaller angles. For a decisive selection more information is required and this may come from the n–p scattering data.

Similar analyses using only s and p waves have been carried out by Klein (1955) and Clementel and Villi (1957). Phillips (1957, private communication) has considered the scattering at 95 Mev in terms of s, p and d waves. From this analysis of the n–p scattering data at 95 Mev (Hillman, Stafford and Whitehead 1956, private communication) and an extrapolation of the p–p data it can be shown that there is a large reduction in the polarization at small angles in p–p scattering due to the nuclear–coulomb interference.

(ii) *Neutron–proton scattering*. Although the purpose of this section is to review high-energy proton–proton scattering, it would be inappropriate to say nothing about neutron–proton scattering. The concept of charge independence of nuclear forces is in itself of great importance and has the consequence that many of the phase shifts are common to both p–p and n–p scattering. More pieces of information are available for distinguishing between sets of phase shifts, but additional phase shifts are required for the n–p interaction which are not present in the p–p interaction owing to the Pauli exclusion principle. By no means as much information is available as for p–p scattering, nor in general is it as accurate. The neutron beams are obtained by charge exchange scattering in light nuclei. In most experiments the neutron beams used have a wide spread in energies (see figures 12 and 13), the neutrons are difficult to detect and polarized beams have until now had quite low polarizations—of the order of 10%. Hence in a double-scattering experiment the asymmetries are much smaller than in the corresponding p–p scattering experiments and in general are more difficult to observe. However, measurements have been made of the free n–p polarization at effective energies of 95 Mev (Hillman, Stafford and Whitehead 1957, private communication), 150 Mev (Roberts, Tinlot and Hafner 1954) and 350 Mev (Siegel, Hartzler and Love 1956). A slightly different experiment has been made at 300 Mev by Chamberlain, Donaldson *et al.* (1954), who used a polarized proton beam and deuterium as a ‘ free ’ neutron target. There is no difference between the two sets of data at 350 and 310 Mev, which are presented in figure 6 as 330 Mev data. Also included are the 95 Mev data. As for p–p polarization, the n–p polarization can be expressed in the form

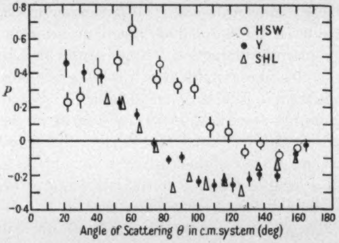


Figure 6. n–p polarization at 95 and 330 Mev. HSW, 95 Mev results of Hillman, Stafford and Whitehead (1956 a) ; Y, 310 Mev results of Ypsilantis (1956) ; SHL, 350 Mev results of Siegel, Hartzler and Love (1956).

$$P\frac{\partial\sigma}{\partial\Omega} = \sin\theta[a + b\cos\theta + c\cos^2\theta + d\cos^3\theta + \ldots]. \qquad \ldots\ldots(4)$$

The constants in this relation are given for the 330 and 95 Mev data as follows :

Effective energy	*a*	*b*	*c*	*d*
330 Mev	−0·34	0·452	1·296	2·928
95 Mev	1·06 ± 0·11	2·34 ± 0·23	0·33 ± 0·39	

So far the 330 Mev n–p data have not been taken into account along with the p–p data in the phase shift analysis reported from Berkeley.

The 95 Mev data have been considered by Phillips. If only s and p waves contribute at 95 Mev, then the polarizations are related in the following way : for p–p scattering $P(\partial\sigma/\partial\Omega) = A\sin 2\theta$, and for n–p scattering $P(\partial\sigma/\partial\Omega) = B\sin\theta + C\sin 2\theta$, with $C = A/4$ if the charge independence hypothesis is valid.

At 140 Mev, $A \sim 0{\cdot}9$ mbn sterad^{-1} and is likely to decrease with decreasing energy, whereas at 95 Mev, $C \sim 1{\cdot}5$ mbn sterad^{-1}. Clearly then either the hypothesis is not valid or higher order waves are required at 95 Mev. It is more reasonable to consider the presence of higher order waves. Phillips is able to obtain agreement with the experimental n–p polarization results by considering s, p and d waves together with a ^{3}S–^{3}D coupling. However, there is a suggestion that the theoretical differential cross section is still a little low at 0° and 180°, indicating that perhaps higher order waves are required. The values he obtains for the constants in equation (4) are $a = 1{\cdot}13$, $b = 2{\cdot}78$ and $c = 0{\cdot}83$. At this energy the only p–p data available are the unpolarized cross sections for θ greater than 40° (Kruse, Teem and Ramsey 1956), but agreement with these is obtained. As it has been necessary to use waves with $L = 2$, with the possibility that higher order waves should be considered, it may be found that some ^3P–^3F interference is still present in the p–p polarization at 95 MEV. Of the solutions proposed by Clementel and Villi (1957) that designated A is preferred by the authors as it gives the best agreement with the p–p results. Restricting the analysis to $L \leqslant 1$ is only valid at energies below 100 MEV and if the n–p scattering data are considered, solution D is to be preferred. In the analysis at 140 MEV, also restricted to s and p waves, the available n–p data were taken into account. Here again only fair agreement is obtained with some of the data. However, with these limitations in mind, these analyses provide a guide if not a solution to the phase shifts.

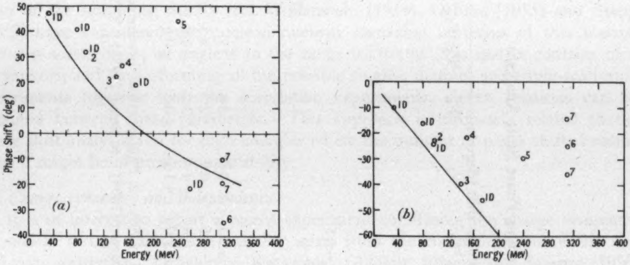


Figure 7. (*a*) Comparison of several determinations of the singlet $^1S_0$ phase shift. The curve is the B solution of Feshbach and Lomon (1956). (*b*) Comparison of several determinations of the triplet $^3P_0$ phase shift. The curve is that given by Feshbach and Lomon (1956). References are given in table 3.

(iii) *Comparison with models.* So far as the present data are concerned, the phase shift analyses have shown that at 300 MEV it is necessary to consider waves up to $L = 4$ or 5 and it may be necessary to go as high as $L = 3$ at 100 MEV. Although the phase shifts have not all been determined without ambiguity, the agreement and variation with energy as displayed in table 3 is such as to suggest that the general features of the scattering are known. It should be possible therefore to make some comparisons with various models of the interaction. As has been stated before, none of the early attempts to explain both the n–p and p–p scattering data in terms of static potentials was very successful. However, it is not ruled out that somewhat different potential models would give better agreement. Up to the present no calculations with other models have been published.†

A charge independent fit to the scattering data has been attempted by Feshbach and Lomon (1956). The model used confined the nuclear forces to a core radius which was dependent upon the state but independent of the energy of the scattering, apart from the $^1S_0$ state. A set of energy independent boundary conditions was used at the core radius for the logarithmic derivative of the wave function. Within this framework the authors were able to obtain reasonable fits to the p–p differential

---

† The models proposed by Signel and Marshak (1957, *Phys. Rev.*, **106**, 832) and by Gammel and Thaler (1957, private communication) gave good agreement with experimental results and predict phase shifts similar to the 1A solution of Clementel and Villi in table 3.

Table 3. Summary of Various Phase Shift Analyses

Ref.	Energy (Mev)	Phase shift (degrees)													
		$^1S_0$	$^1D_2$	$^3P_0$	$^3P_1$	$^3P_2$	$^3F_2$	$^3F_3$	$^3F_4$	$^3P_2$–$^3F_2$	$^1G_4$	$^3F_4$–$^3H_4$	$^3H_4$	$^3H_5$	$^3H_6$
1D	30	47		−10	6	2									
1A	30	47		11·5	−5	4									
1D	95	32		−26	12	2·5									
1A	95	32		20	−12	7									
2	95	32	1	−26	8	6									
3	140	22		−40	10	7									
4	150	26·5	−8	−22	10	8	−6·5								
1D	170	19		−46	12	3									
1A	170	19		21·5	−20·5	9·5									
1D	260	−22		−74	6·5	4									
1A	260	−22		17·5	−25	15·5									
5	240	44		−30	5·4	10									
6	300	−35	1·8	−25	−9	23	−2·3	0·9	3·6	15					
7	300	−19·5	4·3	−36	−12	19	−0·5	0·2	2·5	−9·3	1·3	−1·5	2·1	−1·4	1·6
7	300	−10·1	12·8	−14	−27	16·1	0·8	−4·4	3·2	−0·9	1·0	−1·2	1·4	0·1	1·3
		± 2·5	± 0·7	± 2·1	± 1·3	± 1·0	±1·0	±1·1	±0·5	±0·6					

References : 1A, 1D, Solution A and D of Clementel and Villi (1956) ; 2, Phillips (1956, private communication) ; 3, Author ; 4, Ohnuma and Feldman (1956), Solution F ; 5, Garren (1955) ; 6, Stapp (1955) ; 7, Ypsilantis (1956).

cross section, the n–p total cross sections and differential cross sections for $20^\circ < \theta < 160^\circ$, and approximate fits to the p–p polarizations. For the latter the polarization was assumed to vary as sin 2θ and coulomb–nuclear interference was neglected. Phase shifts for the various states were obtained and the agreement for the $^1S_0$ and $^3P_0$ states with those obtained by others is shown in figures 7 (*a*) and (*b*). As there is good agreement with the values used by Phillips, the Feshbach and Lomon model is expected to fit the n–p polarization data at 95 Mev. The disagreement at the higher energies is perhaps not surprising as the authors state that this model is a rough approximation to the actual interaction.

(*e*) *Scattering matrix.*

Wolfenstein and Askin (1952) and Dalitz (1952) have derived the most general form of the scattering matrix and Wolfenstein (1954), Oehme (1955) and Stapp (1955) have considered the nucleon–nucleon scattering in terms of this matrix. For p–p scattering at an angle θ in the range $0 < \theta < 90^\circ$ this matrix contains nine parameters, and by performing all the possible single-, double- and triple-scattering experiments together with the correlation experiments, eleven relations can be obtained between these parameters. This approach is intimately related to the phase shift analysis, but for high energies where the number of phase shifts become large it might be of greater applicability.

(*f*) *Charge symmetry and independence.*

It is of interest to report a recent experiment confirming the charge symmetry of nuclear forces, namely the equality apart from the coulomb terms of the n–n and p–p scattering. Dzhelepov, Kazarinov, Golovin, Fljagin and Satarov (1956) have studied the quasi-elastic n–p and n–n scattering in a neutron–deuteron scattering at 300 Mev. For large angles the n–d scattering can be regarded as the sum of the n–p and n–n with a small interference term of the order of 15%. Hence by the subtraction of the known n–p scattering from the n–d scattering, it is possible to arrive at a value of the differential n–n cross section. For the angular range $\theta > 50^\circ$ they obtained the value $3{\cdot}7 \pm 0{\cdot}6$ mbn sterad^{-1} which is in excellent agreement with the p–p value. Measurements have also been made at 590 Mev with the same agreement (Dzhelepov, Golovin, Kazarinov and Simonov 1956).

One consequence of the charge independence hypothesis of nuclear force is that certain relations exist between the n–p and p–p scattering cross sections (Feldman 1953). These relations are derived from the isotopic singlet and triplet scattering amplitudes and are triangular in form. The most stringent of them can be expressed as an inequality for the unpolarized differential cross sections, namely

$$\left[\left(\frac{\partial\sigma}{\partial\Omega}\right)_\theta^{\mathrm{pn}}\right]^{1/2} + \left[\left(\frac{\partial\sigma}{\partial\Omega}\right)_\theta^{\mathrm{np}}\right]^{1/2} \geqslant \left[\left(\frac{\partial\sigma}{\partial\Omega}\right)^{\mathrm{pp}}\right]^{1/2}.$$

Since for unpolarized beams

$$\left(\frac{\partial\sigma}{\partial\Omega}\right)_\theta^{\mathrm{np}} = \left(\frac{\partial\sigma}{\partial\Omega}\right)_{\pi-\theta}^{\mathrm{pn}},$$

the inequality at 90° reduces to

$$R = \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}^{\mathrm{pp}} \Big/ \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}^{\mathrm{np}} \leqslant 4.$$

The values of $(\partial\sigma/\partial\Omega)_{90}^{\mathrm{pp}}$ were shown in figure 1 and values for $(\partial\sigma/\partial\Omega)_{90}^{\mathrm{np}}$ were obtained from published values between 40 Mev and 400 Mev (for references, see Randle, Skyrme, Snowden, Taylor, Uridge and Wood 1956). An additional value at 580 Mev has also been reported (Kazarinov and Simonov 1956). Values of R were obtained from these published results and, as is seen in figure 8, there is no violation of the inequality for these energies.

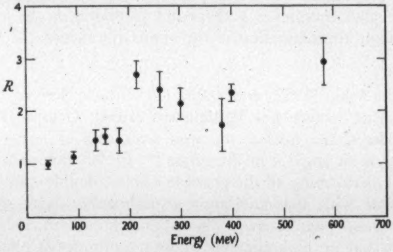


Figure 8. Variation of ratio $R = \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}^{\mathrm{pp}} \Big/ \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}^{\mathrm{np}}$.

The charge independence hypothesis predicts that $R \leqslant 4$.

(g) *Conclusion.*

It might be worth while to indicate the type of data which would now be most useful. As the energy increases, higher order waves enter into the scattering and the number of phase shifts required correspondingly increases. However, the number of possible experiments providing independent pieces of information also increases. Already at 300 Mev the phase shift analysis of the p–p scattering requires fourteen parameters. So far nothing like the same quantity of information is known for the n–p scattering. Whereas at 100 Mev the number of parameters will be smaller, there is the disadvantage that some of the experiments are no longer feasible. The experiments which would be possible are the detailed measurements of polarization and cross section, including the interference region. This latter could provide almost as much information as the triple-scattering experiments. There has been little detailed experimental work in the region 40 to 100 Mev either on p–p or n–p scattering. The analysis of the results in this intermediate energy region would be of value since the validity of the various models is less open to doubt at lower energies. It would also provide a more definite indication of the selection to be made among the numerous fits and sets of phase shifts obtained at the higher energies.

2.3. *Proton–Proton Scattering at Energies above* 400 Mev

(a) *Total cross section.*

As the energy of the measurements has been raised above 400 Mev, it has been found that the character of the p–p interaction changes. The near isotropy in the differential cross section is no longer present and the scattering is peaked at the forward angles. Also the total cross section increases as the production of π mesons becomes energetically possible. Measurements of the total cross section in the range 410 to 660 MeV have been made by Dzhelepov, Moskalev and Medved (1955) by a method similar to that used for the measurements below 400 MeV, except that corrections have to be applied for the mesons that are produced. Fortunately, as the energy is raised, the minimum angle of detection of the particles, $\theta_{\min}$ of equation (2), can be decreased and therefore these corrections become rather small. Similar measurements have been made at 408 MeV (Marshall, Marshall and Nedzel 1953) and at higher energies at Birmingham and Brookhaven. The work of Shapiro, Leavitt and Chen (1954) has been extended to 2600 MeV (Chen, Leavitt and Shapiro 1956), and all these results, together with a smooth curve through the lower energy determinations, are shown in figure 9. The main feature following the almost constant value of 23 mbn between 140 and 300 MeV, is the steep rise to 48 mbn at 800 MeV. The cross section remains at this value up to 1200 MeV and thereafter decreases slightly with further increase in energy. There is some evidence from cosmic-ray work that at 30 000 MeV the cross section is increasing again (Williams 1955). In discussing these variations it is necessary to separate the elastic and inelastic (meson production) components.

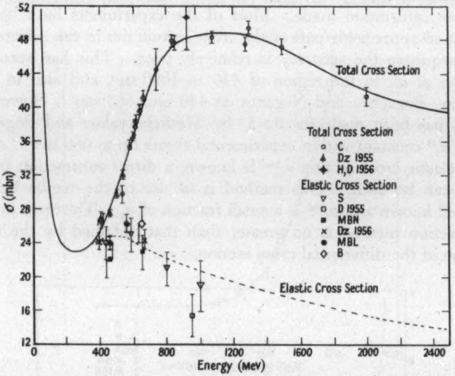


Figure 9. Total and elastic cross section for p–p scattering as a function of energy. Experimental points on total cross section: C, Chen *et al.* (1956); D 1956, Duke *et al.* (1957); H, Hughes, March, Muirhead and Lock (1957); Dz 1955, Dzhelepov, Moskalev and Medved (1955). Experimental points on elastic cross section: S, Smith *et al.* (1955); D 1955, Duke *et al.* (1955); Dz 1956, Dzhelepov, Moskalev *et al.* (1956); MBN, Meshcheryakov, Bogachev and Neganov (1956); B, Batson and Riddiford (1956); MBL, Meshcheryakov, Bogachev, Leksin *et al.* (1956).

(*b*) *Elastic scattering.*

The angular distribution for elastic p–p scattering has been measured by Smith, McReynolds and Snow (1955) for energies between 440 and 1000 MeV, by Batson and Riddiford (1956) at 650 MeV, by Sutton, Fields, Fox, Kane, Mott and Stallwood (1955) at 437 MeV, by Duke, Lock, Marsh, Gibson, McKeague, Hughes and Muirhead (1955) at 950 MeV, by Hartzler and Siegel (1954) at 428 MeV, by Mott, Sutton, Fox and Kane (1953) at 437 MeV, and by a group of workers in Moscow for energies between 410 and 660 Mev (reported by Meshcheryakov, Bogachev and Neganov 1956). Here again the π-meson contamination must be taken into account. From all these experiments it is noticed that, as the energy is increased, the cross sections for large scattering angles become smaller and for small scattering angles larger. The extent of this can be shown in the variation of $(\partial\sigma/\partial\Omega)_{90}$ and the ratio

$$R = \left(\frac{\partial\sigma}{\partial\Omega}\right)_{30} \Big/ \left(\frac{\partial\sigma}{\partial\Omega}\right)_{90}$$

which are plotted in figures 10 and 11.

From these measurements of the elastic differential cross section a value of the total elastic cross section σ_t^{el} can be obtained if the results are extrapolated to zero degrees and an integration made. Most of the experiments have been made for $\theta > 30^\circ$ so that an appreciable part of the cross section lies in the extrapolated region and as a consequence the accuracy is relatively poor. This has been the method used by Smith *et al.* in the region of 440 to 1000 Mev and also in the work of Meshcheryakov, Bogachev and Neganov at 440 and 660 Mev. However, a recent determination has been made for $\theta > 5^\circ$ by Meshcheryakov and Bogachev (1956), who obtained σ_t^{el} constant within experimental errors up to 660 Mev. Alternatively, if the total inelastic cross section σ_t^{inel} is known, a direct subtraction from the total cross section can be used. This method is of use in the region below 600 Mev where σ_t is well known and σ_t^{inel} is a small fraction of σ_t. The error in σ_t^{el} obtained by this subtraction method is no greater than that obtained by the extrapolation and integration of the differential cross section.

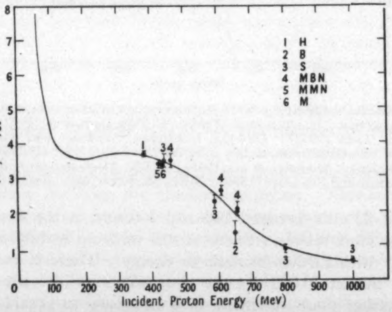


Figure 10. Differential p–p cross section at 90° up to 1000 Mev. Experimental points: 1, Holt (1956); 2, Batson and Riddiford (1956); 3, Smith *et al.* (1955); 4, Meshcheryakov, Bogachev and Neganov (1956); 5, Marshall, Marshall and Nedzel (1953); 6, Mott *et al.* (1953).

(*c*) *Inelastic scattering and meson production.*

The inelastic processes which can take place in this energy interval leading to single meson production are

1 $\mathrm{p+p \rightarrow d+\pi^+}$
2 $\mathrm{p+p \rightarrow p+n+\pi^+}$
3 $\mathrm{p+p \rightarrow p+p+\pi^0}$

The cross section in reaction 1 has been determined between 460 and 660 Mev by Meshcheryakov, Neganov, Bogachev and Sidorov (1955), at 437 Mev by Fields, Fox, Kane, Stallwood and Sutton (1954), at 440 Mev by Rosenfeld (1954), and at 650 Mev by Batson and Riddiford (1956). The reaction 3 has been measured at 480 and 670 Mev (Tyapkin, Kozodaev and Prokoshkin 1955), at 430 Mev (Marshall, Marshall, Nedzel and Warshaw 1952) and at 650 Mev (Batson and Riddiford 1956). Less is known about reaction 2. However, measurements have been made at 650 Mev (Moon 1956), at 660 Mev (Tyapkin and Prokoshkin 1956) and at 410 and 660 Mev (Dzhelepov, Moskalev, Satarov, Golovin and Medved 1956). By suitable extrapolation, values below 650 Mev can be obtained. The cross section for this reaction can be obtained from the measured ratio of reaction 2 to 3. This ratio is approximately the same at 410 and 600 Mev. Adding these three reaction cross sections the total inelastic cross section is obtained, which can then be subtracted from the total cross section to give the total elastic cross section. The values so obtained together with the values from the integrated differential cross section are the same within experimental error. This does not rule out the possibility that between 400 and 800 Mev there is a rather broad and flat resonance in σ_t^{el}. It has been suggested (Chen, Leavitt and Shapiro 1956) that this resonance could be due to an intermediate excited nucleon state. Also the centre of the rise in the σ_t^{inel} curve occurs at a centre-of-mass energy which approximately corresponds to the energy of a peak in the $\pi^+ + p$ cross section. The reader is referred to papers by Rosenfeld (1954), and by Chen, Leavitt and Shapiro (1956) for discussion of these topics.

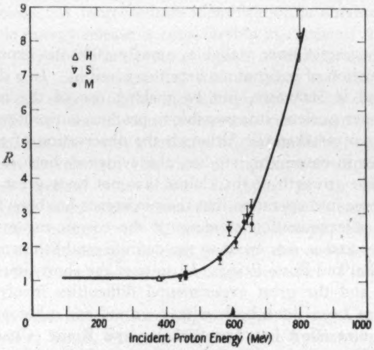


Figure 11. Ratio of $\left(\frac{\partial\sigma}{\partial\Omega}\right)_{30}\Big/\left(\frac{\partial\sigma}{\partial\Omega}\right)_{90} = R$ for p–p scattering as a function of energy. Experimental points : H, Holt (1956) ; S, Smith *et al.* (1955) ; M, Meshcheryakov, Bogachev and Neganov (1956).

The experimental evidence that σ_t^{inel} remains constant at energies above 1000 Mev is deduced from the measured values of σ_t and σ_t^{el}, both of which decrease up to

3000 Mev by about the same amount. Had σ_t^{el} equalled σ_t^{inel} the p–p scattering at these energies might have been considered as that of opaque spheres. The total cross section would then equal $8\pi R^2$ and the corresponding proton radius R would be about $0{\cdot}4 \times 10^{-13}$ cm, a value close to a suggested repulsive core radius but considerably smaller than that obtained in electron–proton scattering (Hofstadter and McAllister 1956). An optical model comparison has been made at these high energies (Serber and Rarita 1955) and is in agreement with a radius between 5 and 10×10^{-14} cm. However, as the energy increases, σ_t^{el} is seen to decrease, showing that the opacity decreases. As the energy is increased beyond 1200 Mev, multiple π-meson production becomes possible, as also does the production of the heavier mesons. For each additional 300 Mev of energy available in the centre of mass, the multiplicity of the production increases by unity.

(*d*) *Antiprotons.*

At the highest energies now available, namely 6000 Mev from the Bevatron at Berkeley, the production of antiprotons becomes possible. In a direct p–p collision the energy required is 5600 Mev, but by making use of the internal motion of nucleons in the copper nucleus, it is possible to produce antiprotons with an incident bombarding energy of 4400 Mev. Although the observation of these particles has been claimed before in cosmic-ray studies, the evidence, both in the frequency of the events and in the strength of the claims, has not been great. It is only since the Bevatron has come into operation that their existence has been finally established. This has led to a re-examination of some of the cosmic-ray events using better analysing techniques and in one event to the definite establishment of an antiproton (Bridge, Caldwell, Pal and Rossi 1956). In spite of the short time that the bevatron has been working and the great experimental difficulties involved, a remarkable amount of work has been done both with counters and photographic emulsions, the latter being a joint effort between Berkeley and Rome (Chamberlain, Chupp, Goldhaber, Segrè, Wiegand, Amaldi, Baroni, Castognoli, Franzinetti and Manfredini 1956).

The first properties of the antiproton to be measured were its mass, charge and stability. Ideally a determination of the production and annihilation processes in hydrogen, spin and parity are also required for the complete identification. As the number produced by the bevatron is rather small, of the order of 20 a day, and as there is a very large background of π^--mesons, elaborate techniques have to be used to detect the antiprotons, and the annihilation and production processes have so far been studied only in heavy elements.

The establishment of the existence of the antiproton among the background of 10^5 π^--mesons involved a selection of momentum by bending magnets, of time of flight over an 80-ft flight path, and finally of velocity by Čerenkov counters (Chamberlain, Segrè, Wiegand and Ypsilantis 1955). The Čerenkov counters were designed to record only the particles within a defined velocity range. From these measurements the mass of the particles was determined on the assumption that they were singly charged. This was borne out by the size of the pulses in one of the counters. The direction of the bending in the magnetic field implied that they were negatively charged. The apparatus was calibrated by using protons and by reversing the direction of the field in the bending magnet. This showed that within the close limits of the apparatus (order of 5%) the antiprotons were of protonic mass. As the antiprotons were travelling with a velocity of 0·75*c*, the time of flight was such that their lifetime must be greater than 10^{-8} second. Since this is longer than that of any heavy unstable particles of comparable mass, it was assumed that the particles were in fact stable.

Measurements have also been made of the annihilation process in heavy elements. One method involved following an antiproton to the end of its track in a photographic emulsion and observing the end-products (Chamberlain, Chupp, Goldhaber *et al.* 1956). The amount of energy visibly released in any star produced should then correspond approximately to the rest mass of the proton if the star is to be identified as an annihilation process. This energy together with that released in neutral particles, assumed to be approximately the same, gives the required total release of twice the rest mass. A more definite identification is obtained from those stars in which the visible energy release is considerably in excess of the rest mass. It is necessary to take only those antiprotons which annihilate near the end of their range in order that they can be identified with certainty and also in order to reduce the uncertainty in the energy of the incident particle which created the star. To achieve these conditions, the incident antiprotons can either be slowed down in a copper absorber so that they will then stop in the emulsion, or alternatively lower energy particles can be selected. In view of the large absorption cross section in copper, the latter method was chosen in spite of an increased π^--meson background. Several stars were observed with visible energy release near or greater than the rest energy, mc^2 (Chamberlain, Chupp, Ekspong *et al.* 1956). The greater part of this energy release goes in the emission of π^--mesons and between two and three are emitted on average per star. Protons and K-mesons are also observed.

The absorption cross section for antiprotons in Be and Cu has been measured (Chamberlain, Keller, Segrè, Steiner, Wiegand and Ypsilantis 1956) by a method similar to that for the p–p total cross section, except that the minimum angle of scattering $\theta_{\min}$ is now about 35° (laboratory angle). This is large enough to include nearly all the diffraction scattering and hence the attenuation is due to inelastic scattering. In order to eliminate the recording of charged secondaries, the inclusion of which would produce a false cross section, a Čerenkov counter was inserted immediately behind the absorber. This was actuated by the charged secondaries but not by the antiprotons, and hence a separation could be made. The attenuation cross section for ordinary protons was also measured with the same apparatus. For both Be and Cu the attenuation cross section for antiprotons was twice that for protons, the cross section in Cu for antiprotons being about $0{\cdot}8 \times 10^{-24}$ cm^2. About 60% of the attenuation cross section was due to annihilation. Estimates from the emulsion work confirm the large attenuation cross section.

There is already a great deal of information which requires explanation, such as the multiplicity of π-meson production and the large absorption cross section for antiprotons. As more work is done, in particular with the annihilation in hydrogen, many of the theories now proposed (Fermi 1950, Belenky and Landau 1956, Sudarshan 1956) will receive a more adequate test. It is of interest to report that before the measurements were made Duerr and Teller (1956) proposed a theory which predicted the large absorption cross section for antiprotons.

§ 3. PROTON–NUCLEAR SCATTERING

3.1. *Introduction*

Where detailed and accurate measurements of the nucleon–nuclear interaction are required protons have certain advantages over neutrons. It is easier to obtain mono-energetic proton beams and to separate elastic from inelastic scattering processes. Furthermore the protons can be counted with 100% efficiency, whereas neutron detection efficiencies are usually less than 10%. However, there is the complication in proton–nuclear scattering of the coulomb interaction. For large scattering angles and high energies this is usually not very important; so that under these conditions the scattering can be regarded as a nucleon–nuclear process and comparisons are then possible between neutron and proton scattering. For small scattering angles interference effects are to be expected and these can be useful in helping to understand the scattering process. Some of the experimental results have been interpreted in terms of an empirical optical model, although recent attempts have been made to obtain the optical model parameters from the nucleon–nucleon data mentioned in § 2.2. (*d*). When the search began for polarization effects in nucleon–nucleon scattering, it was thought that similar effects might be present in proton–nuclear scattering due to quasi-elastic events. A quasi-elastic collision is an elastic collision with a nucleon in the nucleus but is an inelastic process as far as the collision with the nucleus as a whole is concerned. As the greater part of the polarization obtained was found to be due to elastic nuclear scattering, attempts have been made to fit these results using the optical model with a suitable spin–orbit potential added. Recently a clearer separation of elastic and inelastic scattering has been made, and somewhat better agreement has been obtained between the experimental results and the optical model predictions. In the measurement of the inelastic scattering, well-defined groups of protons have been observed corresponding to the excitation of the nucleus to known levels. Hence a more detailed description of the proton–nuclear interaction is required than that given by the optical model. Experiments at high energies are a particularly useful probe of nuclear structure as the wavelength of the incident particles is then less than the inter-nucleon spacing. These experiments have suggested the need for a modification of the independent particle model of the nucleus.

3.2. *Nuclear Transparency and the Goldberger Model*

The mechanism of high-energy collisions was originally proposed by Serber (1947) when it was found *inter alia* that the total cross sections for neutrons decreased below the opaque sphere value of $2\pi R^2$. The nucleus could be regarded as partially transparent. In other words the incident nucleons have a large mean free path in nuclear matter, and can be considered as making collisions only with individual nucleons in the nucleus. These ideas were developed by Goldberger (1948) to predict the number and energies of the outgoing particles. In the initial collision both nucleons might escape, giving a quasi-elastic nucleon–nucleon scattering event, or they could make further collisions, rapidly losing energy and producing a nuclear cascade of particles. Some of these in turn might escape. The prediction about the end-products of the collision was based upon a Monte Carlo calculation in which the collisions of a large number of incident particles were followed and the details of each collision were chosen in a random way. The residual nucleus would be left in a highly excited state and further particles would be 'boiled off' to give the characteristic evaporation spectrum. The detailed appearance of the high-energy end of the spectrum of the emitted particles would depend on the energy of the incident particle, the momentum distribution of the struck nucleons, and the fact that certain collisions would be forbidden by the exclusion principle and by threshold considerations. There would be in addition a group corresponding to the elastically scattered particles.

A typical spectrum of the high-energy neutrons emitted in the forward direction from the proton bombardment of a Be target is shown in figure 12 (Randle, Cassels, Pickavance and Taylor 1953). This was obtained by placing the target inside the cyclotron to intercept a circulating beam of 160 MeV protons, and similar results have been obtained in many laboratories for incident proton energies above 100 MeV.

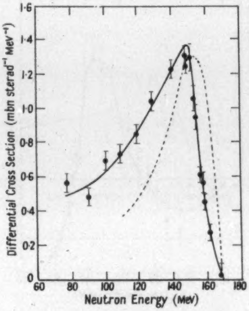


Figure 12. Neutron spectrum in the forward direction from the proton bombardment of beryllium (Randle *et al.* 1953). The broken curve is the computed incident proton spectrum.

Hofmann and Strauch (1953) made measurements for an incident proton energy of 95 MeV and found no peak except for beryllium and even this was attributed to the odd neutron in ^{9}Be. Unfortunately the internal beams of synchrocyclotrons have an appreciable energy spread and, although this can be measured, the spread is such as to vitiate any very detailed deductions about the internal momentum distribution of the nucleons in the nucleus. Nearly all the width of the peak in figure 12 can be attributed to this energy spread in the circulating beam which usually passes through the target several times. Energy and intensity are lost each time so that a tail is built up on the low-energy side. The dotted curve is this computed effective proton spectrum. A similar result is shown for uranium in figure 13 and it is obvious that although there is a decided shift in the peak of the spectrum to lower energies, it is otherwise very similar. This process has been considered by Mandl and Skyrme (1952) in terms of the proton scattering by a neutron in a potential well. They were able to account satisfactorily for the production cross sections.

More direct tests of the Goldberger model have been made by Muirhead and Rosser (1955) for incident protons of 140 MeV, and by Bernardini, Booth and

Lindenbaum (1952) at 340 Mev. The nuclear interactions produced in the elements comprising photographic emulsions were analysed and the model was able to account satisfactorily for the distributions of the emitted particles both in energy and angle, apart from the forward angles. Here the model predicts that the spectrum for the high-energy particles will show a dip since the recoil particle will have a low energy and the Pauli exclusion principle forbids those collisions which lead to occupied momentum states. This was not observed in the photographic plate work nor in a measurement by Snowden (1952) of the angular distribution of neutrons. Some other process is required for the high-energy neutrons emitted in the forward direction. If the system does not have the momentum distribution of a degenerate Fermi gas, assumed in the Goldberger model, then unoccupied low-momentum states are available.

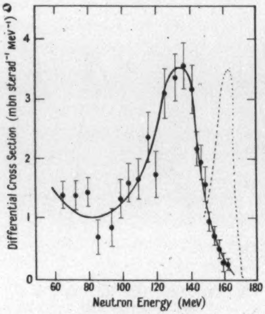


Figure 13. Neutron spectrum in the forward direction from the proton bombardment of uranium (Randle *et al.* 1953). The dotted curve is the computed incident proton spectrum.

3.3. *Inelastic Scattering Groups*

The Goldberger model is a statistical model and predicts a continuous spectrum of emitted particles since the detailed nature of the nuclear structure does not enter into the calculation. Careful measurements of the high-energy protons inelastically scattered from carbon at 96 Mev (Strauch and Titus 1956) have shown the presence of groups of protons corresponding to the excitation of the nucleus to well-known low levels. Similar results have been obtained at 185 Mev (Tyrén and Maris 1957). Carbon was chosen as the first level is 4·4 Mev above the ground state, so that with incident beams having an energy spread of 1 or 2 Mev it was possible to separate the elastic and inelastic events. Measurements were made by Strauch and Titus at angles greater than the first diffraction minimum, since for these angles the elastic and inelastic events occurred in comparable numbers. The detection method was not able to separate clearly the relatively smaller inelastic contribution at smaller angles. The peaks occur at energies in the proton–carbon centre-of-mass system independent of the angle of observation and therefore the collision has occurred with the nucleus as a whole. The separation of the peaks from the elastic peak suggests that levels at 4·4, 9·6 and 20·8 Mev were being excited.

The known level at 7·7 Mev did not seem to be strongly excited. No diffraction maxima or minima were observed for these groups.

In the measurements with 185 Mev protons (Tyrén and Maris 1957), the various groups were separated in a magnetic analyser which had an energy resolution of 2 Mev. The incident beam had an energy spread of something better than 1 Mev. It was therefore possible to separate clearly the 4·4 Mev group from the elastically scattered protons and make measurements at angles less than the first diffraction minimum. In figures 14 (*a*) and (*b*) the spectra shown are for scattering angles of 11° and 29°. The groups corresponding to the 4·4 and 9·6 Mev levels show a dip in the cross section at the forward scattering angles but the 7·7 Mev group appears to have no such dip. In contrast the 15 and 20 Mev groups appear to have larger cross sections at smaller scattering angles. The measurements at angles beyond the diffraction minimum for other light elements, which have been made by Strauch (1956, private communication) at 95 Mev have shown similar effects to carbon in that certain low-lying levels are

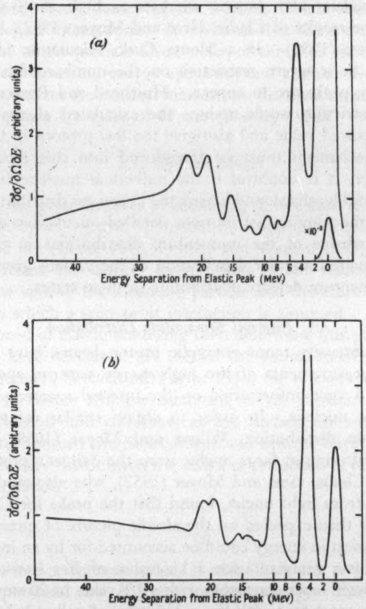


Figure 14. Inelastic scattering from carbon for 185 Mev protons (Tyrén and Maris 1957). (*a*) at 11° ; (*b*) at 29°.

strongly excited and have differential cross sections comparable to those for elastic scattering. However, for heavier elements, where the spacing of the low-lying levels is closer, the elastic scattering seems always to predominate.

Strauch suggests that the groups may be due to individual nucleon–nucleon collisions in the nucleus but that the individual nucleons do not behave as free particles. On the individual particle model the collision can be imagined as raising the struck nucleon into a higher energy state. Since the mean free path is large for the incident particle, it leaves the nucleus without making a further collision. To explain the 4·4 MeV group at large scattering angles requires the presence of nucleons in the nucleus with kinetic energies as high as 40 MeV, which is not improbable from the results of Cladis, Hess and Moyer (1952), Wilcox and Moyer (1955) and of Selove (1956). In a Monte Carlo calculation for the Goldberger model, there must be a severe restriction on the number of collisions permitted if these high-energy peaks are to appear. Muirhead and Rosser (1955) have estimated that this restriction would reduce the calculated absorption cross section well below the measured value and also give too few protons in the inelastic peaks. Therefore some mechanism must be introduced into this model to weight the collisions. However, it is doubtful if the individual nucleon in the nucleus can be regarded as free and localized, and having the Fermi gas distribution of momentum. In order to predict the angular distributions detailed calculations are required which would involve knowledge of the momentum distributions in the various states. Conversely the measured angular distribution of the various groups could be used to estimate the momentum density distribution in these states.

3.4. *Internal Momentum Distribution*

Now that high-intensity mono-energetic proton beams have been extracted, a repetition of the measurements of the high-energy neutron spectra from proton bombardment would yield information on the internal momentum distribution of the neutrons in the nucleus. In order to obtain similar information about the corresponding proton distribution, Wilcox and Moyer (1955) have studied the quasi-elastic p–p scattering at large angles from the 340 MeV proton bombardment of light elements. Cladis, Hess and Moyer (1952), who also studied the scattering of 340 MeV protons from light nuclei, found that the peaks for C and O were at a slightly lower energy than expected on the simple picture of quasi-elastic scattering but that the distribution in energy could be accounted for by an internal momentum distribution given by a gaussian with a 1/e value of 16 ± 3 MeV. The results of Wilcox and Moyer were more consistent with a 1/e value of 20 MeV. The spectrum from deuterium was consistent with the momentum distribution obtained from the Hulthén wave function.

Further information about the momentum distribution of the nucleons in the nucleus has been obtained from measurements of pick-up deuterons (Selove 1956, Chew and Goldberger 1950). In the 95 MeV measurements of Selove the energy spectrum of the deuterons from the proton bombardment of carbon showed that the pick-up process could leave the residual ^{11}C nucleus in a low-lying excited state. The momentum distribution of that particular state governs the angular distribution of the picked-up deuteron. It was found that at all angles the majority of the deuterons corresponded to the formation of ^{11}C in its ground state, and possibly in a 4·5 MeV level. Analysis of the angular distribution of the main group showed that the density of the higher momentum components was greater than that given by a gaussian with a 1/e value of 16 MeV. In fact these components were too great to be accounted for by the independent particle model and this suggests that the interaction of a nucleon in a carbon nucleus involves a deep short-range well in addition to the smoothed potential well of the independent particle model. It had been difficult previously to understand the great success of the smoothed potential in view of the very strong short-range forces in nucleon–nucleon interactions. Brueckner, Eden and Francis (1955) have derived these high-momentum components from a many-body theory of the nucleus which takes into account the strong nucleon–nucleon interactions. It is shown that these components have only small effects on the single-particle observables of the shell model. The threshold for production, yield function, energy spectrum and angular distribution in meson production also depend on the internal momentum distribution and can be accounted for by a gaussian with a 1/e value of 19 MeV (Henley 1952).

3.5. *Optical Model*

(*a*) *Introduction.*

In considering the gross features of scattering of nucleons it is possible to represent the interaction by an ' optical ' potential in which the detailed nuclear structure has been smoothed out (Bethe 1940, Francis and Watson 1953). Elastic scattering distributions determine the density distributions of the nucleons. The potential may depend on the spin of the incoming particle and therefore optical models have been developed in which a spin–orbit interaction is assumed. Experiments on the polarization produced in elastic scattering then determine this potential.

Many authors have developed the idea of the optical model. From the optical point of view (Fernbach, Serber and Taylor 1949) the nucleus is regarded as having a complex refractive index and the scattering of the incident particles as that of light waves. Reflection and refraction at the surface are neglected. It is seen from the similarity between the Schrödinger and optical wave equations that for a uniform nuclear density distribution the index of refraction is related to the complex potential by the equation

$$\eta^2 - 1 = \frac{V}{E} + i\frac{W}{E}$$

where η is the complex refractive index, V and W are the real and imaginary parts of the ' optical ' potential and E is the incident energy. Since for high energies $\eta^2 - 1$ is small and the wavelength of the incident particle is small compared with inter-nucleon distances, it is possible to consider the transmitted rays as suffering a change in phase and amplitude depending on the distance travelled. The application of Huygens' principle to these rays then gives the scattering amplitude. (The similar electromagnetic case has been worked out by van de Hulst (1946).) This model is not likely to be valid for large scattering angles owing to the simplifying assumptions made. However, it has been shown by Shapiro (1955) that the model gives results not very different from an exact phase shift calculation even out to quite large angles. It can be further shown that the forward scattering cross section is not sensitive to the shape of the potential.

The imaginary part of the potential, and therefore of the refractive index, controls the absorption. The coefficient of absorption K on a particle picture can be put as $K = 3A\sigma/4\pi R^3$ where σ is the effective cross section for scattering by a nucleon in the nucleus of radius R and atomic weight A. This σ can be expressed in terms of the individual nucleon–nucleon cross sections and for an incident proton is given by

$$A\sigma = Z\sigma_{\mathrm{pp}}\alpha_{\mathrm{pp}} + N\sigma_{\mathrm{pn}}\alpha_{\mathrm{pn}}$$

where Z and N are the numbers of protons and neutrons respectively, σ_{pp} and σ_{pn} the respective nucleon–proton cross sections, and α is a parameter less than unity allowing for the Pauli exclusion of those collisions which lead to an occupied state. If the refractive index is put as

$$\eta = \frac{k+k_1}{k} + i\frac{K}{2k}$$

where k is the wave number outside the nucleus and $k+k_1$ inside, then K/k approximately equals W/E. Goldberger (1948) calculated the effective cross section σ for 90 MeV neutrons assuming that the nucleons in the nucleus had a momentum distribution corresponding to a degenerate Fermi gas. The measurements mentioned above have shown that higher momentum components are present and the system may not be degenerate. Therefore the Goldberger estimate of α is likely to be low. The total absorption cross section is given by the simple relation

$$\sigma_{\mathrm{a}} = \pi R^2\{1-(1-[1+2KR]\exp[-2KR])/2K^2R^2\}.$$

Measurements of σ_{a} can therefore be used to determine R if the calculated value of K is used or alternatively some preconceived notion of R can be used and an empirical value of K obtained.

The scattering amplitude of $f(\theta)$ obtained by Shapiro in a more exact treatment is

$$f(\theta) = i\int_0^\infty \mathrm{J}_0(2k\rho\sin\tfrac{1}{2}\theta)\,[1-\exp(-Ks+2ik_1s)]\,\rho d\rho$$

where ρ is the impact parameter, $2s$ the path length of the transmitted ray such that $\rho^2+s^2 = R^2$; J_0 is the Bessel function of zero order. This formula for $f(\theta)$ differs by a factor i from that obtained in the direct optical derivation (Fernbach, Serber and Taylor 1949).

(*b*) *Total and differential cross sections.*

Several authors have considered the neutron cross sections in terms of the square well optical model (Taylor 1953, Shapiro 1955) and found empirical values for the various parameters. The potentials were assumed to vary smoothly with energy but were considered independent of atomic weight. The radius R was assumed to be independent of the energy. Remarkably good fits were obtained to the data over an energy range of 40 to 400 MeV. If account can be taken of the coulomb interaction, it is tempting to apply the same model to the proton scattering results. Measurements of the total cross sections for protons have been made at a number of energies. It is not possible to measure these in 'good' geometry because of the large coulomb scattering at small angles, and the measurements have therefore to be made in 'poor' geometry and extrapolated to 'good' geometry conditions. For the same reasons the measurements have been restricted to light elements. Since the coulomb and nuclear scattering is coherent, strong interference effects will occur in the region where the respective amplitudes of scattering are comparable.

Determinations have been made of the total nuclear cross section of light elements for protons at 134 MEV (Taylor and Wood, unpublished), at 408 MEV (Marshall, Marshall and Nedzel 1953), at 208 and 315 MEV (de Carvalho 1954) and at 860 MEV (Chen, Leavitt and Shapiro 1955). The measurements are made of the cross section for the scattering beyond angles greater than some given angle θ, and if several values of θ are chosen, the extrapolation to 'good' geometry consists of a continuation of the variation of σ with θ to $\theta = 0$. Alternatively, estimates can be made of the corrections to be applied for the diffraction and coulomb scattering, and these can be obtained from the theoretical angular distributions. At the angles at which most measurements have been made there is considerable destructive nuclear–coulomb interference, as can be seen in figure 15. Hence in order to extract the purely nuclear part of the cross section, it is necessary to know the extent and magnitude of this interference. For the 860 MEV data this interference effect was probably small and was neglected, but the errors on the determination were increased. For the 134 MEV data the interference was obtained by fitting the measured angular distribution for protons to an optical model prediction which took account of the coulomb interaction. The interference and pure coulomb scattering was then taken as the difference between this optical model with and without the modification for the coulomb scattering. The nuclear potentials used in fitting the angular distribution were found to be identical with those given by Taylor (1953) for the same neutron energy. Values obtained for the total nuclear cross section in all the experiments were the same, within the errors, as those from

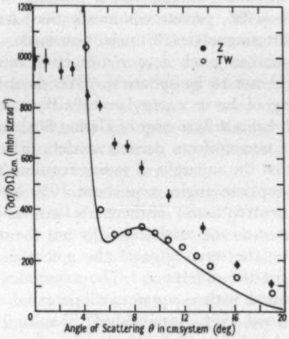


Figure 15. Angular distribution of elastically scattered protons and neutrons from C. Z, van Zyl *et al.* (1956), 136 MEV; TW, Taylor and Wood (1956, unpublished), 142 MEV. The curve is a theoretical optical model calculation for protons.

the determination for neutrons. There was a slight difference for Be at 134 MeV which could be attributed to the difference of the n–p and p–p cross section arising from the odd neutron.

Measurements of the elastic scattering differential cross section have also been made by Richardson, Ball, Leith and Moyer (1952) at 340 MeV and Strauch and Titus (1956) at 95 MeV. Although detailed variations in the angular distribution for 340 MeV protons were observed, there were no pronounced diffraction maxima and minima, especially for the light elements. Strauch and Titus were able to find these for carbon only by a very careful separation of the elastic events. The non-appearance of deep diffraction minima is partly due to the transparency of the nucleus, but even allowing for this the theoretical minima were still much too low. Further factors which can raise these minima are the presence of a diffuse edge to the nucleus, of a spin–orbit potential and a possibly non-spherical shape. On this latter point Strauch (1956, private communication) has compared the diffraction patterns at large scattering angles of nuclei near lead. Lead, which is assumed to be a spherical nucleus, has much deeper diffraction minima than neighbouring nuclei which are thought not to be spherical. Woods and Saxon (1954) in their analysis of the scattering of lower energy protons have found the need to use a radial dependence which has a diffuse edge. Gatha, Shah and Patel (1954) analysed the 340 MeV data with a non-uniform density model. Shapiro (1955) in analysing the 95 MeV data found that the square-well radius required to reproduce the diffraction minima at the appropriate angles was about 10% less than that obtained by Taylor (1953) from the neutron data. If there is a long tail on the nuclear potential, this will give a contribution to the total scattering but the diffraction minimum will be mainly determined by the steep edge of the potential. Therefore it may not be difficult to account for the difference. The experimental results in the interference region can be fitted with a square-well potential and a suitable coulomb potential (Taylor and Wood 1956, unpublished, Shapiro 1955). As the measurements in this region can be made with high accuracy, they can be used to determine accurately the magnitude of the real part of the potential for any assumed shape.

(*c*) *Absorption cross sections and nuclear radii.*

The complications associated with the coulomb interaction do not apply to measurements of the absorption cross section as these can be made in 'poor' geometry where the coulomb interaction contributes little to the scattering. Cassels and Lawson (1954) have measured the absorption cross section for 134 MeV protons for a number of elements. Assuming the value of K calculated on the Goldberger model, they obtained a value of R which was consistent with that obtained from low-energy neutron scattering data. This is given approximately by $R = (0{\cdot}57 + 1{\cdot}30A^{1/3}) \times 10^{-13}$ cm and is close to the value obtained by Taylor (1953) from an analysis of the neutron data between 40 and 400 MeV. Measurements of the absorption cross section for protons have also been made at 860 MeV (Chen, Leavitt and Shapiro 1955) and at 290 MeV (Millburn, Birnbaum, Crandall and Schecter 1954). In terms of the same optical model, the higher energy determinations seem to require a radius of $R = 1{\cdot}25A^{1/3} \times 10^{-13}$ cm.

Since the distribution of the neutrons in the nucleus may extend further than that of the protons, different values of the overall size might be obtained from measurements with protons and neutrons between 100 and 400 MeV. At these energies the proton–neutron cross section is larger than the proton–proton cross section, so that one of the internal distributions mainly governs the scattering. For instance, for incident protons the overall size determined from the absorption cross section will more nearly correspond to the extent of the neutron distribution. It has been suggested (Johnson and Teller 1954) that this is greater than that of the protons. Using both a square-well and a tapered density optical model, and making a correction for coulomb repulsion, Voss and Wilson (1955) compared the radii determined from neutron and proton absorption measurements. Since the radii for heavy elements determined in the proton measurements were greater than in the neutron measurements by about 10%, they attributed this difference to a greater extent of the neutron distribution in the heavy nuclei. At the higher energies the two nucleon–nucleon cross sections are more nearly equal and it is not then possible to make the above distinction. However, the analysis of the 1400 MeV neutron measurements (Coor, Hill, Hornyak, Smith and Snow 1955) has led to much smaller radii which are more in agreement with the charge distribution radii determined from electron scattering (Hofstadter, Hahn, Knudsen and McIntyre 1954). Williams (1955), analysing the same data, assumed a shape for the variation of the nuclear density and potential similar to that obtained for the charge distribution in the electron scattering experiments. The extents of the two distributions were then found to be very similar.

For the analysis of proton scattering at energies close to 20 MeV, Woods and Saxon (1954) used a potential of the form $V = V_0/\{1+\exp[(R-R_0)/a]\}$ where a determines the thickness of the diffuse edge. Excellent agreement with experimental results was obtained, but with radii greater by about 10% than the charge distribution radii. It has been suggested that the low-energy data are more sensitive to the range of nuclear forces and that it is possible to assume the same radii for charge and matter distribution. The difference between the extent of the nuclear potential (determined in proton and neutron scattering experiments) and charge distribution (determined in electron scattering experiments) is then attributed to the effect of a long-range nuclear force (Drell 1956). Even so the neutron distribution may have a slightly longer tail and surface thickness than the proton distribution (Wilets 1956). The nuclear potential is estimated to extend $0{\cdot}7\times10^{-13}$ cm beyond the nuclear matter.

3.6. *Polarization*

(*a*) *Introduction.*

As already mentioned, attempts to explain the nucleon–nucleon scattering data in terms of static potentials led to the consideration of non-central terms in the interaction and to a search for polarization effects. It was difficult to insert hydrogen targets into cyclotrons, so that Be and C targets were originally used to produce the partially polarized beams. It was thought that at energies above 100 MeV the quasi-elastic nucleon–nucleon collisions in the nucleus would produce about the same amount of polarization as in the free nucleon–nucleon collisions. Measurements were made therefore of the azimuthal asymmetry in a double scattering just as in the p–p experiments. As the angle of the first scattering was such that there was usually an appreciable amount of elastic as well as the quasi-elastic scattering, it was necessary to know the asymmetry as a function of the energy of the scattered particle. This was required so that the magnitude of the polarization could be determined from the asymmetry recorded in a double-scattering experiment, both scatterings of which were to be identical apart from energy loss. Most measurements were made with a range counter telescope which recorded the number of particles with range, and therefore energy, greater than some given value. It was soon found that the measured asymmetry, and therefore the polarization, was greatest for the highest energy particles, i.e. those elastically rather than those quasi-elastically scattered.

(*b*) *Elastic scattering and spin–orbit coupling.*

Fermi (1955) has suggested that the strong spin–orbit coupling which had been postulated in the shell model (Mayer 1950) might be operative in the elastic scattering of nucleons. This would manifest itself as a difference in the energy of the interaction according as the spin and orbital angular momentum were parallel or antiparallel. Since the range of nuclear forces is smaller than nuclear radii, this spin–orbit interaction would only be operative near the nuclear surface (on average the forces in the interior will balance). Thus the effective spin–orbit potential would depend on the gradient of the nuclear density as well as the orientation of the spin and angular momentum and could be represented as

$$V_s = -\frac{a}{r}\frac{\partial V}{\partial r}\mathbf{l}.\mathbf{s}, \qquad \ldots\ldots(8)$$

where V is usually taken as the real part of the nuclear potential $V+iW$, $\mathbf{l}$ and $\mathbf{s}$ are vectors representing the orbital angular momentum and spin respectively and a is the coupling constant. This assumes that the radial dependence of the potential is the same as that of the nuclear density. The value of the coupling constant required to explain the experimental results is some twenty to thirty times greater than that expected from a Thomas precession but is about equal to that used in the shell model. As at these energies the collisions of the incident particles are with individual nucleons in the nucleus and as the potentials are a smoothed representation of the interaction in such collisions, it is perhaps more appropriate to look upon the spin dependence in the nucleon–nucleus interaction as a consequence of the established spin dependence in the nucleon–nucleon interaction (Francis and Watson 1953, Tamor 1955, Takeda and Watson 1955). On this basis and using some simplifying assumptions, Fernbach, Heckrotte and Lepore (1955) were able to justify the form of V_s given in equation (8).

(*c*) *Variation with energy.*

A calculation in Born approximation gives the following polarization (Fermi 1954, Malenka 1954),

$$P(\theta) = \frac{2k^2 a(W/V)\sin\theta}{1+(W/V)^2+k^4a^2\sin^2\theta}. \qquad \ldots\ldots(9)$$

Although the Born approximation will not be valid except for small scattering angles and light elements, equation (9) will display the main features of the polarization. Thus within the limits of the approximation $P(\theta)$ is independent of well shape and nuclear radius. If W or V is zero, then there is no polarization. Further, as the energy is increased above 50 MeV, V decreases but W increases slightly (Taylor 1953) and therefore $P(\theta)$ is expected to increase. It is indeed found that the polarization increases rapidly at energies above 50 MeV.

The maximum polarization is given by (Malenka 1954, Levintov 1956)

$$P_{\max} = \frac{W/V}{[1+(W/V)^2]^{1/2}} \qquad \ldots\ldots(10)$$

and occurs at an angle $\theta_{\max}$ given by

$$\sin\theta_{\max} = \frac{[1+(W/V)^2]^{1/2}}{ak^2}.$$

Values of W and V can be taken from the optical model used to fit the neutron scattering results (Taylor 1953, Shapiro 1955) and the value of $P_{\max}$ so obtained is shown in figure 16. The experimental values for carbon included for comparison have been obtained by Dickson, Rose and Salter (1956, unpublished) for energies below 150 MeV, by Strauch (1955) for energies between 141 and 314 MeV, by Chamberlain, Segrè, Tripp, Wiegand and Ypsilantis (1956) at 289 and 315 MeV and by Chesnut, Hafner and Roberts (1956) at 220 MeV. There is seen to be fair agreement but there is the suggestion that $P_{\max}$ may not be as great above 250 MeV as at 150 MeV. Approximate agreement is also obtained for $\theta_{\max}$. However, here there is less certainty about the value to be taken, as the experimental results either vary little with θ near $\theta_{\max}$ or can have included inelastic scattering.

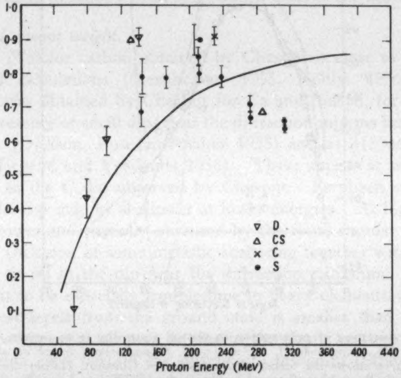


Figure 16. Variation of maximum polarization in proton elastic scattering from carbon as a function of energy. Experimental points: D, Dickson, Rose and Salter (1956, unpublished); CS, Chamberlain, Segrè *et al.* (1956); C, Chesnut, Hafner and Roberts (1956); S, Strauch (1955). The curve is that given by equation (10) when values of V and W are taken from Taylor (1953).

(*d*) *Variation with scattering angle.*

More exact calculations have been made using the WKB approximation (Fernbach, Heckrotte and Lepore 1955), which is equivalent to a calculation using the optical model of Fernbach, Serber and Taylor (1949). This yields somewhat different results from the Born approximation for large angles. In the angular region near the first diffraction minimum $P(\theta)$ shows a violent dip for a central square-well potential. Since the effective potential for $\mathbf{s}$ and $\mathbf{l}$ parallel is slightly different from that when they are antiparallel, the diffraction minima will occur at slightly different angles. Thus for a square well and opaque nucleus when measurements are made on one side at the angle corresponding to the diffraction minimum, all the particles of an incident completely polarized beam will be scattered to the opposite side, giving $P(\theta) = \pm 1$. At a slightly different angle corresponding to the diffraction minimum on the other side $P(\theta) = \mp 1$. Even if potential forms are used which round off the edge of the nucleus, for instance the Woods and Saxon potential (1954), and allowance is made for transparency, this dip is not completely eliminated provided W/V is not small. As the energy is decreased and W/V decreases the dip tends to get smaller.

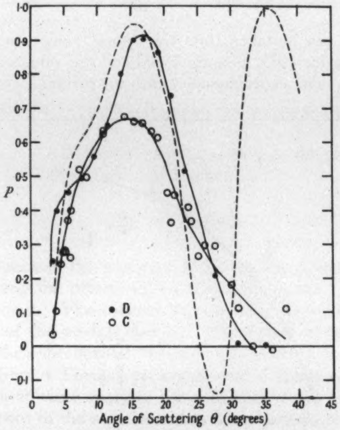


Figure 17. Angular variation of polarization in proton scattering from carbon. D, Dickson, Rose and Salter (1956, unpublished), 137 Mev ; C, Chamberlain, Segrè *et al.* (1956), 300 Mev ; the broken curve shows the trend of the results of Chesnut, Hafner and Roberts (1956) at 220 Mev. Angles have been changed by the factor $(E/220)^{\frac{1}{2}}$ to bring the diffraction minimum to the same angle for the three energy determinations.

Of the experimental results for carbon, only that of Chesnut, Hafner and Roberts (1956) at 220 Mev shows a very deep dip. In spite of a 12 Mev spread in energy of the incident beam, careful measurements were made to separate inelastic and elastic scattering. The trend of the result obtained by Chesnut is depicted in figure 17 where it is seen that the purely elastic scattering does have this dip. The measurements at Harwell and Berkeley, although approximately reproducing the 220 MeV results at angles up to the dip, showed zero polarization beyond. Chesnut's measurements also showed that in the region of the diffraction minimum where $P(\theta)$ for the elastic scattering is essentially zero, $P(\theta)$ for the inelastic scattering is very high. Strauch and Titus (1956) and Tyrén and Maris (1957) have shown that the inelastic scattering from the 4·4 MeV level and the elastic scattering are comparable in intensity near the diffraction minimum (figure 14 (*b*)). Thus a small amount of inelastic scattering can produce a large change or 'dilution' in the polarization. Although this can account in some measure for the fact that the 220 MeV result is the only one to show a very deep dip for carbon, it would be difficult to attribute to it the non-appearance of the second $P_{\max}$ at the other energies shown in figure 17 (Dickson, Rose and Salter 1956, unpublished, Chamberlain, Segrè, Tripp, Wiegand and Ypsilantis 1956)†. However, all this does emphasize that great care must be taken if the relation $\Sigma = P^2$ is used in two supposedly identical scatterings (equation (3)), since in many experiments the angle of the first scattering used to produce a polarized beam is close to the diffraction minimum.

(*e*) *Variation with atomic weight.*

The form of $P(\theta)$ for carbon obtained by Chesnut is close to that obtained in most theoretical calculations (Sternheimer 1955, Köhler 1956, for instance). A similar result was obtained by Chesnut for Ca and indeed, for the elements Al and above, the presence of small dips near the diffraction minima have been observed both at 140 MeV (Dickson, Rose and Salter 1955) and at 315 MeV (Chamberlain, Segrè, Tripp, Wiegand and Ypsilantis 1956). These appear at smaller angles but are not as deep as the C dip observed by Chesnut. Fernbach *et al.* (1955) have suggested that the dip may be shallower at lower energies. At higher energies the dip may be narrower and partially obscured by the finite angular resolution of the determination. Inclusion of some inelastic scattering together with a finite angular resolution tends to fill in the dip near the diffraction minimum. This 'dilution' might be thought to be especially troublesome in heavy elements where the separation of the lowest levels from the ground state is smaller than in carbon. The fact that the dips appear at all may be due to the elastic scattering predominating for heavy elements even beyond the diffraction minimum (Strauch 1956, private communication). The general trend of these results is a decrease of the first maximum in P as the atomic weight is increased and the appearance of second and higher maxima of P at larger angles (figures 18 and 19). This decrease in the first maximum of P may be attributed to coulomb effects (Fernbach *et al.* 1955). Neighbouring elements seem to have very similar polarization patterns, which suggests that proton-nuclear polarization is independent of nuclear spin.

Sternheimer (1955) has made some detailed calculations of the polarization using a realistic shape for the potential and including the coulomb interaction. A Woods–Saxon potential was used with parameters which for an optical model calculation gave agreement with experimentally determined absorption and total cross sections. A spin–orbit potential of form given by equation (8) with V_s about 1 MeV was added to the central potential. This value of V_s is about the same as that used in the shell model of the nucleus. The calculation for carbon does not predict a very pronounced dip. A similar calculation for iron predicted a very deep dip near the diffraction minimum, which was not observed. Some of this discrepancy is certainly due to experimental 'dilution', but on the other hand there is no certain form of the potential to be used. Indeed, the spin–orbit potential is likely to depend on the derivative of the nuclear density (Fermi 1955) rather than on that of the nuclear potential which may be different (see § 3.5 (*b*)). Exact measurements in the region of the diffraction minimum might provide a guide to the radial dependence of the spin–orbit potential (Skyrme 1956, private communication).

---

† Agreement with the polarization measurements at 220 MeV both for elastic and inelastic scattering from the 4·4 MeV level is obtained in work at Uppsala at 155 MeV (ALPHONSE, R., JOHANSSON, A., and TIBELL, G., 1957, *Nucl. Phys.*, in the press). Also a clearer separation at 137 MeV gives better agreement than indicated in figure 17.

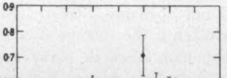


Figure 18. Angular variation of the polarization in proton scattering from aluminium at 287 MeV. The experimental points are those of Chamberlain, Segrè *et al.* (1956).

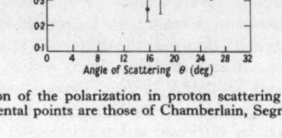


Figure 19. Combined results of the angular variation of the polarization in proton scattering from silver, cadmium and indium. The experimental points are those of Dickson, Rose and Salter (1956, unpublished) and the curve that of Köhler (1956). The finite angular resolution can account for almost all the difference between 12° and 16°.

(*f*) *Coulomb interference.*

As has been indicated in figure 15 for the unpolarized cross section, coulomb effects are particularly important at small angles and it is to be expected that comparable effects will be present in the polarization (Wilson 1955). These effects are apparent in figure 20, which shows the results of Dickson, Rose and Salter (1956, unpublished) for scattering from C for several energies. On the Born approximation, which for carbon at these angles is reasonably valid, $P(\theta)$ is approximately proportional to θ. From the chain line through the results it is seen that in the interference region there is a distinct departure from this variation with θ. If the dotted extrapolation is assumed to represent the neutron polarization, then it is found that the ratio of the polarization is approximately equal to the inverse ratio of the corresponding cross sections. However, the coulomb interaction has the effect of reducing the first maximum in P and increasing the angular width of the dip over that for neutrons (Fernbach *et al.* 1955). Therefore this extrapolation does not truly represent the neutron polarization and the above relationship between cross sections and polarizations is not exactly satisfied.

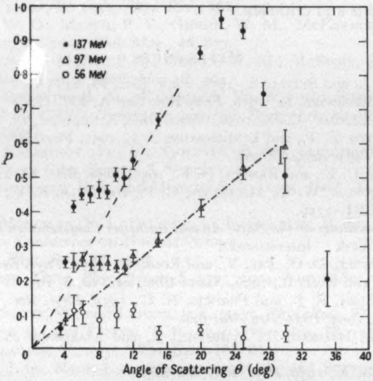


Figure 20. Angular variation of polarization in proton scattering from carbon at 137, 97 and 56 MeV, showing effects of nuclear coulomb interference at small angles (Dickson, Rose and Salter 1956, unpublished).

(*g*) *Imaginary part of spin–orbit potential.*

So far the spin–orbit potential has been assumed to be real and proportional to the derivative of the real part of the central potential. However, it is known that the inelastic scattering is polarized and it therefore seems necessary to include an imaginary part. Heckrotte (1956) found it necessary to include this in order to achieve agreement with the measurements at small angles of the polarization of 315 MeV protons scattered by carbon (Chamberlain, Segrè, Tripp, Wiegand and Ypsilantis 1956). The calculation included relativistic coulomb corrections which are important for small angles. It was found that to obtain agreement with the experimental results the imaginary part had to be equal in magnitude but of opposite sign to the real part of the nuclear spin–orbit potential, whose sign had to be the same as in the shell model of the nucleus. Thus it was possible to confirm the sign of the polarization from this small-angle region. The existence of the imaginary part of the spin–orbit potential has been deduced by Riesenfeld and Watson (1956).

The optical model parameters, i.e. the scattering potentials, were expressed in terms of the nucleon–nucleon phase shifts as given by Feshbach and Lomon (1956). It was shown that the complex spin–orbit potential was similar to that used by Heckrotte (1956) and that the central potential was considerably smaller than that derived by neglecting spin–orbit coupling (Taylor 1953). A calculation in Born approximation of the polarization in proton scattering from carbon was broadly in agreement with experimental results at all energies.

Acknowledgments

I would like to express my thanks to the many authors who allowed me to see the results of their work before publication. Thanks are also due to many associates for useful discussions and criticism.

References

Baskir, E., Tinlot, J., and Spital, S., 1955, *Phys. Rev.*, **99**, 639 (A).
Batson, A. P., and Riddiford, L., 1956, *Proc. Roy. Soc.* A, **237**, 175.
Belenky, S. Z., and Landau, L. D., 1956, *Nuovo Cim.*, **3**, Suppl. 1, 15.
Bernardini, G., Booth, E. T., and Lindenbaum, S. J., 1952, *Phys. Rev.*, **88**, 1017.
Bethe, H. A., 1940, *Phys. Rev.*, **57**, 567.
Birge, R. W., Kruse, U. E., and Ramsey, N. F., 1951, *Phys. Rev.*, **83**, 274.
Birnbaum, W., Crandall, W. E., Millburn, G. P., and Pyle, R. V., 1954, University of California, UCRL–2756.
Breit, G., 1956, *Proceedings of the Sixth Annual Rochester Conference on High Energy Nuclear Physics* (New York: Interscience).
Bridge, H. S., Caldwell, D. O., Pal, Y., and Rossi, B., 1956, *Phys. Rev.*, **102**, 930.
Brinkworth, M. J., and Rose, B., 1956, *Nuovo Cim.*, Ser. 10, **3**, 195.
Brueckner, K. A., Eden, R. J., and Francis, N. C., 1955, *Phys. Rev.*, **98**, 1445.
de Carvalho, H. G., 1954, *Phys. Rev.*, **96**, 398.
de Carvalho, H. G., Heiberg, E., Marshall, J., and Marshall, L., 1954, *Phys. Rev.*, **94**, 1796.
Case, K. M., and Pais, A., 1950, *Phys. Rev.*, **80**, 203.
Cassels, J. M., 1956, *Proc. Phys. Soc.* A, **69**, 495.
Cassels, J. M., and Lawson, J. D., 1954, *Proc. Phys. Soc.* A, **67**, 125.
Cassels, J. M., Pickavance, T. G., and Stafford, G. H., 1952, *Proc. Phys. Soc.* A, **214**, 262.
Chamberlain, O., Chupp, W. W., Ekspong, A. G., Goldhaber, G., Goldhaber, S., Lofgren, E. J., Segrè, E., Wiegand, C., Amaldi, E., Baroni, G., Castagnoli, C., Franzinetti, C., and Manfredini, A., 1956, *Phys. Rev.*, **102**, 921.
Chamberlain, O., Chupp, W. W., Goldhaber, G., Segrè, E., Wiegand, C., Amaldi, E., Baroni, G., Castagnoli, C., Franzinetti, C., and Manfredini, A., 1956, *Phys. Rev.*, **101**, 909.
Chamberlain, O., Donaldson, R., Segrè, E., Tripp, R., Wiegand, C., and Ypsilantis, T., 1954, *Phys. Rev.*, **95**, 850.
Chamberlain, O., and Garrison, J. D., 1954, *Phys. Rev.*, **95**, 1349.
Chamberlain, O., Keller, D. V., Segrè, E., Steiner, A. M., Wiegand, C., and Ypsilantis, T., 1956, *Phys. Rev.*, **102**, 1637.
Chamberlain, O., Pettengill, G., Segrè, E., and Wiegand, C., 1954, *Phys. Rev.*, **93**, 1424.
Chamberlain, O., Segrè, E., Tripp, R., Wiegand, C., and Ypsilantis, T., 1954, *Phys. Rev.*, **93**, 1430.
Chamberlain, O., Segrè, E., Tripp, R., Wiegand, C., and Ypsilantis, T., 1956, *Phys. Rev.*, **102**, 1659.
Chamberlain, O., Segrè, E., and Wiegand, C., 1951, *Phys. Rev.*, **83**, 923.
Chamberlain, O., Segrè, E., Wiegand, C., and Ypsilantis, T., 1955, *Phys. Rev.*, **100**, 947.

Chen, F. F., Leavitt, C. P., and Shapiro, A. M., 1955, *Phys. Rev.*, **99**, 857; 1956, *Phys. Rev.*, **103**, 211.

Chesnut, W. G., Hafner, E. M., and Roberts, A., 1956, *Phys. Rev.*, **104**, 449.

Chew, G. F., and Goldberger, M. L., 1950, *Phys. Rev.*, **77**, 470.

Christian, R. S., 1952, *Rep. Progr. Phys.*, **15**, 68 (London: Physical Society).

Christian, R. S., and Noyes, H. P., 1950, *Phys. Rev.*, **79**, 85.

Cladis, J. B., Hess, W. N., and Moyer, B. J., 1952, *Phys. Rev.*, **87**, 425.

Clementel, E., and Villi, C., 1957, *Nuovo Cim.*, in the press.

Coor, T., Hill, D. A., Hornyak, W. F., Smith, L. W., and Snow, G., 1955, *Phys. Rev.*, **98**, 1369.

Dalitz, R. H., 1952, *Proc. Phys. Soc.* A, **65**, 175.

Dickson, J. M., Rose, B., and Salter, D. C., 1955, *Proc. Phys. Soc.* A, **68**, 361.

Dickson, J. M., and Salter, D. C., 1954, *Nature, Lond.*, **173**, 946.

Drell, S. D., 1956, *Phys. Rev.*, **100**, 97.

Duerr, H. P., and Teller, E., 1956, *Phys. Rev.*, **101**, 494.

Duke, P. J., Lock, W. O., March, P. V., Gibson, W. M., McKeague, R., Hughes, I. S., and Muirhead, H., 1955, *Phil. Mag.*, **46**, 877.

Duke, P. J., Lock, W. O., March, P. V., Gibson, W. M., McEwen, J. G., Hughes, I. S., and Muirhead, H., 1957, *Phil. Mag.*, **2**, 204.

Dzhelepov, V. P., Golovin, B. M., Kazarinov, Yu. M., and Simonov, Yu. M., 1956, *Proceedings of the CERN Symposium on High Energy Accelerators and Pion Physics* (Geneva: CERN).

Dzhelepov, V. P., Kazarinov, Yu. M., Golovin, B. M., Fljagin, V. B., and Satarov, V. I., 1956, *Nuovo Cim.*, Ser. 10, 3, Suppl. 1, 61.

Dzhelepov, V. P., Moskalev, V. I., and Medved, S. V., 1955, *Dokl. Akad. Nauk., S.S.S.R.*, **104**, 380.

Dzhelepov, V. P., Moskalev, V. I., Satarov, V. I., Golovin, B. M., and Medved, S. V., 1956, Moscow Conference on Nuclear Physics.

Feldman, D., 1953, *Phys. Rev.*, **89**, 1159.

Fermi, E., 1950, *Progr. Theor. Phys., Japan*, **5**, 570; 1954, *Nuovo Cim.*, **11**, 407; 1955, *Ibid.*, **2**, Suppl., 17.

Fernbach, S., Heckrotte, W., and Lepore, J. V., 1955, *Phys. Rev.*, **97**, 1059.

Fernbach, S., Serber, R., and Taylor, T. B., 1949, *Phys. Rev.*, **75**, 1352.

Feshbach, H., and Lomon, E., 1956, *Phys. Rev.*, **102**, 891.

Fields, T. H., Fox, J. G., Kane, J. A., Stallwood, R. A., and Sutton, R. B., 1954, *Phys. Rev.*, **95**, 638 (A).

Fischer, D., and Baldwin, J., 1955, *Phys. Rev.*, **100**, 1445.

Fischer, D., and Goldhaber, G., 1954, *Phys. Rev.*, **95**, 1350.

Francis, N. C., and Watson, K. M., 1953, *Phys. Rev.*, **92**, 291.

Garren, A., 1955, Carnegie Institute of Technology, *Report NYO* 7102; 1956, *Phys. Rev.*, **101**, 419.

Garrison, J. D., 1954, *Thesis*, University of California, UCRL–2659.

Gatha, K. M., Shah, G. Z., and Patel, N. J., 1954, *Proc. Phys. Soc.* A, **67**, 773.

Gell-Mann, M., and Watson, K., 1954, *Ann. Rev. Nucl. Sci.*, **4**, 219.

Goldberger, M. L., 1948, *Phys. Rev.*, **74**, 1269.

Goldfarb, L. J. B., and Feldman, D., 1952, *Phys. Rev.*, **88**, 1099.

Hartzler, A. J., and Siegel, R. T., 1954, *Phys. Rev.*, **95**, 185.

Heckrotte, W., 1956, *Phys. Rev.*, **101**, 1406.

Heiberg, E., Kruse, U., Marshall, J., Marshall, L., and Solmitz, F., 1955, *Phys. Rev.*, **97**, 250.

Henley, E. M., 1952, *Phys. Rev.*, **85**, 204.

Heusinkveld, M., and Freier, G., 1952, *Phys. Rev.*, **85**, 80.

Hillman, P., Stafford, G. H., and Whitehead, C., 1957, *Nuovo Cim.*, in the press.

Hofmann, J. A., and Strauch, K., 1953, *Phys. Rev.*, **90**, 449.

Hofstadter, R., Hahn, B., Knudsen, A. W., and McIntyre, J. A., 1954, *Phys. Rev.*, **95**, 512.

Hofstadter, R., and McAllister, R. W., 1955, *Phys. Rev.*, **98**, 217.

Holt, J. R., 1956, *Proceedings of the Sixth Annual Rochester Conference on High Energy Nuclear Physics* (New York : Interscience).

Hughes, I. S., March, P. V., Muirhead, H., and Lock, W. O., 1957, *Phil. Mag.*, **2**, 215.

van de Hulst, H. C., 1956, *Rech. Astr. Obs. Utrecht*, **11**, Pt. 1.

Jastrow, R., 1951, *Phys. Rev.*, **81**, 165.

Johnson, M. H., and Teller, E., 1954, *Phys. Rev.*, **93**, 357.

Kane, J. A., Stallwood, R. A., Sutton, R. B., Fields, T. H., and Fox, J. G., 1954, *Phys. Rev.*, **95**, 1694.

Kazarinov, Yu. M., and Simonov, Yu. N., 1956, Moscow Conference on Nuclear Physics.

Klein, C. A., 1955, *Nuovo Cim.*, **1**, 581.

Köhler, H. S., 1956, *Nuclear Physics*, **1**, 433 (Amsterdam : North-Holland).

Kruse, U. E., Teem, J. M., and Ramsey, N. F., 1956, *Phys. Rev.*, **101**, 1079.

Levintov, I. I., 1956, *Dokl. Akad. Nauk., S.S.S.R.*, **107**, 240.

Malenka, B. J., 1954, *Phys. Rev.*, **95**, 522.

Mandl, F., and Skyrme, T. H. R., 1952, *Proc. Phys. Soc.* A, **65**, 101.

Marshall, J., and Marshall, L., 1955, *Phys. Rev.*, **98**, 1398.

Marshall, J., Marshall, L., and de Carvalho, H. G., 1954, *Phys. Rev.*, **93**, 1431.

Marshall, J., Marshall, L., and Nedzel, V. A., 1953, *Phys. Rev.*, **91**, 767.

Marshall, J., Marshall, L., Nedzel, V. A., and Warshaw, S. D., 1952, *Phys. Rev.*, **88**, 632.

Mayer, M. G., 1950, *Phys. Rev.*, **78**, 16, 22.

Meshcheryakov, M. G., and Bogachev, N. P., 1956, Moscow Conference on Nuclear Physics.

Meshcheryakov, M. G., Bogachev, N. P., Leksin, G. A., Neganov, B. S., and Piskarev, E. V., 1956, *Proceedings of the CERN Symposium on Physics High Energy Accelerators and Pion* (Geneva : CERN).

Meshcheryakov, M. G., Bogachev, N. P., and Neganov, B. S., 1956, *Nuovo Cim.*, **3**, Suppl. 1, 119.

Meshcheryakov, M. G., Neganov, B. S., Bogachev, N. P., and Sidorov, U. M., 1955, *Dokl. Akad. Nauk., S.S.S.R.*, **100**, 673, 677.

Millburn, G. P., Birnbaum, W., Crandall, W. E., and Schecter, L., 1954, *Phys. Rev.*, **95**, 1268.

Moon, P. B., 1956, *Proceedings of the Sixth Annual Rochester Conference on High Energy Nuclear Physics* (New York : Interscience).

Mott, N. F., 1932, *Proc. Roy. Soc.* A, **135**, 429.

Mott, W. E., Sutton, R. B., Fox, J. G., and Kane, J. A., 1953, *Phys. Rev.*, **90**, 712.

Muirhead, H., and Rosser, W. G. V., 1955, *Phil. Mag.*, **46**, 652.

Oehme, R., 1955, *Phys. Rev.*, **98**, 147, 216.

Ohnuma, S., and Feldman, D., 1956, *Phys. Rev.*, **102**, 1641.

Oxley, C. L., Cartwright, W. F., and Rouvina, J., 1954, *Phys. Rev.*, **93**, 806.

Oxley, C. L., and Schamberger, R. D., 1952, *Phys. Rev.*, **85**, 416.

Pettengill, G. H., 1954, *Thesis*, University of California, UCRL–2808.

Randle, T. C., Cassels, J. M., Pickavance, T. G., and Taylor, A. E., 1953, *Phil. Mag.*, **44**, 425.

Randle, T. C., Skyrme, D., Snowden, M., Taylor, A. E., Uridge, F., and Wood, E., 1956, *Proc. Phys. Soc.* A, **69**, 760.

Richardson, R. E., Ball, W. P., Leith, C. E., and Moyer, B. J., 1952, *Phys. Rev.*, **86**, 29.

Riesenfeld, W. B., and Watson, K. M., 1956, *Phys. Rev.*, **102**, 1157.

Roberts, A., Tinlot, J., and Hafner, E. M., 1954, *Phys. Rev.*, **95**, 1099.

Rosenfeld, A. H., 1954, *Phys. Rev.*, **95**, 638 (A) ; **96**, 130, 139.

Selove, W., 1956, *Phys. Rev.*, **101**, 231.

Serber, R., 1947, *Phys. Rev.*, **72**, 1114.

Serber, R., and Rarita, W., 1955, *Phys. Rev.*, **99**, 629 (A).

Shapiro, A. M., Leavitt, C. P., and Chen, F. F., 1954, *Phys. Rev.*, **95**, 663 (A).

Shapiro, I. I., 1955, *Thesis*, Harvard University.

Siegel, R. T., Hartzler, A. J., and Love, W. A., 1956, *Phys. Rev.*, **101**, 838.

Smith, L. W., McReynolds, A. W., and Snow, G., 1955, *Phys. Rev.*, **97**, 1186.

Snowden, M., 1952, *Phil. Mag.*, **43**, 285.

STAPP, H. P., 1955, University of California, UCRL 3098.

STERNHEIMER, R. M., 1955, *Phys. Rev.*, **100**, 886.

STRAUCH, K., 1955, University of California, UCRL 3211.

STRAUCH, K., and TITUS, W. F., 1956, *Phys. Rev.*, **103**, 200.

SUDARSHAN, G., 1956, *Phys. Rev.*, **103**, 777.

SUTTON, R. B., FIELDS, T. H., FOX, J. G., KANE, J. A., MOTT, W. E., and STALLWOOD, R. A., 1955, *Phys. Rev.*, **97**, 783.

SWANSON, D. R., 1953, *Phys. Rev.*, **89**, 740.

TAKEDA, G., and WATSON, K. M., 1955, *Phys. Rev.*, **97**, 1336.

TAMOR, S., 1955, *Phys. Rev.*, **97**, 1077.

TAYLOR, A. E., and WOOD, E., 1956 a, *Proc. Phys. Soc.* A, **69**, 645 ; 1956 b, *Proceedings of the Sixth Annual Rochester Conference on High Energy Nuclear Physics* (New York : Interscience).

TAYLOR, T. B., 1953, *Phys. Rev.*, **92**, 831.

THALER, R. M., BENGSTON, J., and BREIT, G., 1954, *Phys. Rev.*, **94**, 683, 679.

TOWLER, O. A., 1951, *Phys. Rev.*, **84**, 1262.

TRIPP, R. D., 1955, University of California, UCRL 3186.

TYAPKIN, A. A., KOZODAEV, M. S., PROKOSHKIN, YU. D., 1955, *Dokl. Akad. Nauk., S.S.S.R.*, **100**, 689.

TYAPKIN, A. A., and PROKOSHKIN, YU. D., 1956, Moscow Conference on Nuclear Physics.

TYRÉN, H., and MARIS, TH. A. J., 1957, *Nucl. Phys.*, **3**, 35.

VOSS, R. G. P., and WILSON, R., 1955, *Phys. Rev.*, **99**, 1056.

WILCOX, J. M., and MOYER, B. J., 1955, *Phys. Rev.*, **99**, 875.

WILETS, L., 1956, *Phys. Rev.*, **101**, 1805.

WILLIAMS, R. W., 1955, *Phys. Rev.*, **98**, 1393.

WILSON, R., 1955, *Phil. Mag.*, **46**, 769.

WOLFENSTEIN, L., 1949, *Phys. Rev.*, **76**, 541 ; 1954, *Ibid.*, **96**, 1654 ; 1956, *Ann. Rev. Nucl. Sci.*, in the press.

WOLFENSTEIN, L., and ASKIN, J., 1952, *Phys. Rev.*, **85**, 947.

WOODS, R. D., and SAXON, D. S., 1954, *Phys. Rev.*, **95**, 577.

WOUTERS, L., 1951, *Phys. Rev.*, **84**, 1069.

YPSILANTIS, T. J., 1955, *Thesis*, University of California, UCRL 3047.

YPSILANTIS, T., 1956, *Proceedings of the Sixth Annual Rochester Conference on High Energy Nuclear Physics* (New York : Interscience).

YPSILANTIS, T., WIEGAND, C., TRIPP, R., SEGRÈ, E., and CHAMBERLAIN, O., 1955, *Phys. Rev.*, **98**, 840.

VAN ZYL, C. P., VOSS, R. G. P., and WILSON, R., 1956, *Phil. Mag.*, **1**, 1003.

THEORIES OF COLLECTIVE BEHAVIOUR

By D. TER HAAR
The Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	130
§ 2. General considerations	138
§ 3. Sound waves	143
§ 4. Plasma oscillations	148
§ 5. Spin waves	149
§ 6. Collective behaviour of nuclei	153
§ 7. Liquid helium	155
Acknowledgments	159
References	159

Abstract. A general discussion of the problems raised by systems of many interacting particles is given in the introduction. In the second section general methods are discussed for dealing with these problems, and special attention is paid to the method developed by Tomonaga. This method is applied to the theories of sound waves (§ 3), plasma oscillations (§ 4), nuclear surface modes (§ 6), and liquid helium (§ 7). A brief discussion is also given of some aspects of spin wave theory (§ 5).

---

§ 1. INTRODUCTION

In recent years a large amount of work has been done on many-particle problems. We shall try to discuss some of the attempts made to solve these problems, to show the relations between the various theories, and to point out some of the remaining problems.

Let us start by summarizing briefly the present state of theoretical physics, as this shows very clearly the reasons for this gradually increasing interest in many-body systems. There are, and have been for quite a long time, subjects in physics where the state of theoretical understanding is satisfactory. We may mention optics, thermodynamics, electrodynamics, and probably hydrodynamics. On the other hand, quantum mechanics, nuclear physics, solid state physics, and the theory of liquids are not in such a happy state. It will be noted that we have split theoretical physics roughly into two categories, the one dealing with phenomenological, macroscopic descriptions of physical systems, the other endeavouring to describe systems in microscopic terms, starting from first principles. There are many unsolved problems as to what are the first principles to be used—we may mention the struggles of field theorists to understand (or even classify) the galaxy of 'elementary' particles. However, even if we think that we know from what principles to start, the resulting problem is of such a complexity that one must have recourse to a great many simplifications and one is no longer certain that the final results of lengthy computations bear much relation to the properties of systems which they purport to describe.

The reason why in solid state physics, for instance, one is so often led to extremely complicated equations is that one is essentially always presented with a many-body problem. Here we meet with another dichotomy, one which exists in those parts of theoretical physics dealing with microscopic behaviour. As long as the problem can be simplified so as to lead to essentially a single-particle problem, progress can be made relatively easily. Although at first sight this may look like a very serious restriction, it is in fact surprising how many problems can be reduced to single-particle (or perhaps rather single quasi-particle) problems and we shall, indeed, discuss several instances of this in the present review. The simplest one-particle problem is that of classical mechanics when one considers the motion of a particle in a given field of force. A similar problem arises when one considers the behaviour of two interacting particles. In that case one can neglect the motion of the centre of gravity and one is left with the motion of one mass point in a given field of force. The centre of gravity follows also a one-particle orbit, in the absence of external forces a straight line. It is perhaps instructive to consider this simple case in somewhat more detail since this two-body system already shows several features of the many-body problem. Originally we were considering two interacting particles, the orbits of which may be described by the Cartesian coordinate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. It is, however, more convenient to describe the system in terms of a *collective coordinate* $\mathbf{X}\,[=(m_1\mathbf{x}_1+m_2\mathbf{x}_2)/(m_1+m_2)]$, where m_1 and m_2 are the masses of the two particles, and a *particle coordinate* $\mathbf{x}(=\mathbf{x}_1-\mathbf{x}_2)$. The collective coordinate is here the position of the centre of gravity and the particle coordinate the relative position of the two particles with respect to one another.

We shall take a relatively general case where the system is present in a homogeneous gravitational field corresponding to a force $m\mathbf{g}$ on a particle of mass m and where the force exerted by particle 1 on particle 2, which by Newton's third law is equal to the negative of the force exerted by particle 2 on particle 1, is $\mathbf{F}$. The equations of motion in terms of $\mathbf{x}_1$ and $\mathbf{x}_2$ are

$$m_1\ddot{\mathbf{x}}_1 = m_1\mathbf{g}+\mathbf{F}, \quad m_2\ddot{\mathbf{x}}_2 = m_2\mathbf{g}-\mathbf{F}. \qquad \ldots\ldots(1.1)$$

In terms of $\mathbf{x}$ and $\mathbf{X}$ we have instead the equations of motion

$$[M = m_1+m_2, \quad \mu = m_1m_2/(m_1+m_2)]$$

$$M\ddot{\mathbf{X}} = M\mathbf{g}, \qquad \ldots\ldots(1.2a)$$

$$\mu\ddot{\mathbf{x}} = \mathbf{F}. \qquad \ldots\ldots(1.2b)$$

Equation (1.2*a*) describes the collective motion: the system as a whole moves in the gravitational field. Equation (1.2*b*) describes the individual motion: the resultant orbit is that of a particle with mass μ moving under the influence of a force $\mathbf{F}$. We are, indeed, dealing with the behaviour of a quasi-particle; instead of a particle of mass m_1 or m_2 we find a 'particle' with mass μ!

As soon as we go one step further and consider a three-body system, we are faced with difficulties. The three-body problem has been considered in great detail in classical mechanics, especially by astronomers (see, for instance, Poincaré 1892), but an exact general solution is not known and one has recourse to approximate solutions. The situation is, of course, even worse when we consider systems of more constituents in classical mechanics.

One might hope that the situation would improve when we are dealing with systems containing so many particles that statistical methods can be used, that is, where the number of particles is so large that fluctuations are negligibly small.

Indeed, systems containing many particles can often be treated more satisfactorily than systems containing only a few particles. The reason for this is that if there are only a few particles present it is necessary to consider the orbits and behaviour of the constituents in detail, while only the average behaviour is of interest in a many-particle system. The problem is thus reduced to that of the average behaviour. In the case where the particles are practically non-interacting—so that we are dealing with a perfect gas—the average behaviour is easily ascertained and statistical mechanics can deal satisfactorily with this case. The case of systems which are so dilute that only two-body encounters need be considered can also be treated easily. One is essentially dealing with the second virial coefficient and both quantum mechanics and classical mechanics will give usable expressions for the deviations from the perfect gas behaviour as long as three-body and more complicated collisions can be neglected. If, however, three-body collisions are important, the theory is no longer able to give a satisfactory general expression for the third virial coefficients. In classical statistics, one can obtain an expression for the third virial coefficient in terms of the interaction potential between two particles (see, for instance, ter Haar 1954 a, p. 179), which has been used by de Boer and Michels (1939) and by Montroll and Mayer (1941) to compare theoretical and experimental values. It must be realized, however, that this expression is derived by making the assumption that the intermolecular forces are additive, that is, that it is only necessary, in principle, to know the potential energy of a pair of particles in order to evaluate the third (and higher) virial coefficients. Although at moderate densities this assumption is unlikely to lead to serious mistakes, there is no reason to expect it to be correct and, indeed, there are indications (McGinnies and Jansen 1956) that it is incorrect. The situation is even more complicated in quantum statistics and up to the present nobody has attempted to evaluate the quantum mechanical third virial coefficient even assuming additivity, except for the trivial case of the perfect Bose–Einstein and Fermi–Dirac gases.

When the density of a gas is increased higher virial coefficients come into play and it should be clear from our considerations that statistical mechanics will be hard put to describe such systems, unless simplifying assumptions are made. Such systems have been studied in great detail in recent years and although much progress has been made by Yvon, Kirkwood, Born, and collaborators (see, for instance, Yvon 1934, Kirkwood 1946 and later papers in the *Journal of Chemical Physics*, Born and Green 1949, de Boer 1949, Green 1952, Hirschfelder, Curtiss and Bird 1954), the discussion at the recent Statistical Mechanics Conference (Brussels 1957) made it abundantly clear that much is still to be done in this field, not the least a clarification of the fundamental ideas on which present theories are based. In the case of dense gases the behaviour of conglomerations of several particles which are acting collectively is important as can be seen from the fact that higher virial coefficients play a role. We are clearly in the field of many-body theories. The situation is even worse when we come to the problem of condensation. Mayer's theory (1937, Mayer and Mayer 1940) was elucidated by Kahn and Uhlenbeck (1938, Kahn 1938) and mathematically put in its final form by Yang and Lee (1952), but there is still no physical understanding of the condensation process and we are far from being able to predict under what circumstances actual condensation will set in.

Although the Yvon–Kirkwood–Born–Green theories are also intended to cover the liquid state (see, for instance, Green 1952), it is well known that liquids are notoriously badly understood, except perhaps that fascinating liquid He II which is slowly being understood. The reason for this may well be that we can consider it at the absolute zero. Indeed, Feynman's work has shown us how to treat a quantum Bose–Einstein liquid at and near absolute zero. We shall return to liquid helium presently.

When we come to solids the situation simplifies to a large extent. In fact, it is often stated that statistical mechanics is able to account satisfactorily for the behaviour of crystals. This is true only up to a point. If we may treat the crystal as a perfect one, and if we may use the Debye approximation to describe the crystal vibrations (for an account of the Debye theory of crystal vibrations see, for instance, Seitz 1940) we can, indeed, calculate the specific heat due to the crystal vibrations and we find the T^3 behaviour which to a first approximation is found experimentally. Similarly, statistical mechanics is able to describe qualitatively order–disorder transitions in substitutional alloys (see, for instance, Nix and Shockley 1938, ter Haar 1954 a, chap. XII), at any rate for simple cubic and body-centred cubic structures. Finally, many transport properties of solids, especially of metals, can be understood qualitatively. For metals one uses as a first approximation the Sommerfeld free electron (or rather quasi-free electron) theory (Sommerfeld 1928, Sommerfeld and Bethe 1933). This theory gives a fair qualitative picture, but it is very difficult to improve it quantitatively although a great deal of work has been done, notably by Wilson and collaborators (see Wilson 1953).

We can easily understand why it is not possible to give quantitative theories of the phenomena which we mentioned a moment ago. It is far more difficult to see why the approximate theories are such good approximations. One of the aims of the present article is to give reasons for this. Broadly speaking it is because the problems can be reduced to one-(quasi-)particle problems. In the case of the lattice vibrations the quasi-particles are the phonons, and the 'free electrons' of the Sommerfeld theory are really wave packets moving through the metal. For many purposes it is possible to treat these wave packets as negative charges surrounded by a screening positive charge cloud as has been shown by Bohm and Pines (1951, 1953, Pines and Bohm 1952, Pines 1953, 1954, 1955, 1956; see also Raimes 1957). The case of the substitutional alloys is more complicated, but it is probably correct to say that the possibility of constructing approximate theories—and for the two-dimensional Ising model a rigorous theory (Onsager 1944)—is connected with the fact that the basis assumption is made of essentially nearest neighbour interaction only so that the problem is reduced to a (usually extremely complicated) purely combinatorial problem.

In the case of crystal vibrations we are clearly dealing with collective behaviour of all the particles in the system. In the case of the conduction electrons on the one hand we have the collective behaviour of all the other conduction electrons which will produce the screening cloud, and on the other hand the periodic structure of the lattice—that is, the collective behaviour of the nuclei which form such a lattice—will lead to energy bands and thus to the occurrence of an 'effective mass'. Another typical example of a quasi-particle is the occurrence of 'holes' which can to a very large extent be treated as particles, though obviously one can never collect them in a box. Other cases in solid state physics where clearly the constituents of the system act collectively are ferromagnetism (and antiferromagnetism), superconductivity, and the occurrence of excitons and polarons.

Let us briefly consider the phenomenon of superconductivity. Recently Fröhlich (1950) has developed a theory of superconductivity which was successful especially in predicting the so-called isotope effect. This theory is essentially a one-electron theory, working as it does with the concept of a Fermi sphere in momentum space. It has been shown by several people (for instance, Schafroth 1951, 1952, Fröhlich 1954) that the theory is not consistent, but it remains all the same remarkable that such a theory could be developed. In this particular case it has not yet been possible to find the reason for the success of the Fröhlich theory, but it looks as if investigations along the lines suggested by Tisza (1950) might be useful.

There are other physical phenomena where collective behaviour is such that it leads to an independent (quasi-)particle model. The first case is that of the shell model of the nucleus. The second case is that of the mass- and charge-normalization of the electron. The third case is that of the magnetic moments of the proton and the neutron. Let us briefly consider these cases. The first case is a very clear one. As in the case of the shell model of the atom it is possible to assign to the different particles one-particle orbits and to fill up the various shells. The physical properties of the final system depend mainly on the orbits of the particles outside the closed shells. In the atomic case the one-particle orbits are found by using the Hartree–Fock method (Hartree 1928, Fock 1930 ; for an account see, for instance, Schiff 1949, chap. XI). That is, one assumes the action of the other electrons to be such as to produce a background field in which the electron considered moves. Recent work by Brueckner, Eden, Bethe, and collaborators (see Bethe 1956, Eden 1956) has shown that, even though nuclear forces are strong and have a short range, the same procedure can be used. The Pauli principle plays an important role, as had been predicted by Weisskopf (1951).

That the ' bare ' mass of an electron is different from the ' experimental ' mass, a fact predicted by the classical electron theory (Lorentz 1909), was shown most dramatically by the Lamb–Retherford (1947) experiment. This led to a renewed consideration of renormalization theories, the earliest of which was the one of Kramers (1938, 1957). Here the collective action of the quanta in the radiative electromagnetic field leads to an additional contribution to the inertial mass of the electron, just as the collective action of the quanta in the meson field surrounding the neutron leads to its non-vanishing magnetic moment. A similar effect exists in the case of the proton and of the electron, although in the latter case the g-factor is only slightly different from the ' bare ' value.

Up to now we have discussed how in systems of many degrees of freedom many of these degrees of freedom can often combine to produce certain effects, and we would thus be tempted to classify all these phenomena as collective effects. This is certainly correct, but it is not in conformity with the usual nomenclature, and we ought to specify in somewhat more detail what we shall call collective modes and what we shall call individual modes. We must, however, immediately emphasize that this distinction is artificial and often arbitrary. This is, perhaps, most easily demonstrated by looking at the case of the lattice vibrations. In that case the behaviour of the N interacting particles ($3N$ degrees of freedom) in the lattice is described by means of the $3N$ normal modes. Each mode in turn will correspond to a number of quanta (phonons). At first sight we would call the lattice vibrations a typical case of collective behaviour, since each particle is involved in a given normal mode. However, if we look at it from the point of view of the phonons, we are dealing with a (nearly independent) particle system. It thus depends largely on the method of approach and the problem considered whether or not to consider lattice vibrations to fall under the heading ' collective behaviour ' or under the heading ' individual behaviour '. The phonons are a typical example of 'quasi-particles ', and the lattice vibrations can to a large extent be treated as a perfect gas of phonons. Such a theory is, for instance, able to give a fair kinetic theory for the lattice heat conduction. In order that a quasi-particle approach can be useful, it is necessary that the quasi-particles are independent, or at least nearly independent. In that case the main properties of interest are their energy levels and their statistics. Of course, in order that we can apply statistics to the perfect gas of quasi-particles it is necessary that the number of quasi-particles present in the system is large. We see thus why the plasmons although otherwise being typical quasi-particles cannot under normal circumstances lead to a perfect gas type of theory of plasma oscillations. Under normal circumstances their density is too small for statistical methods to be of use.

With the introduction of the concept of quasi-particles we have in principle solved the question of why independent particle theories are often so successful. The answer is that the collective action of the system as a whole often leads to excitations which behave as quasi-particles, and that those quasi-particles in many respects resemble the constituent particles of the system. For many properties the main point is the statistics. This is especially the case for degenerate Fermi–Dirac systems where most quantities depend only on the values of various functions of the energy at the surface of the Fermi sphere so that the actual shape of the energy spectrum is of secondary importance. This aspect of the quasi-particle theories has been developed by Bontsch-Brujewitsch (1955) to whom we also refer for an extensive bibliography, especially of the Russian literature. We may also refer to articles by Wonssowski (1954) and Pekar (1954 b, also 1954 a) for references to Russian articles on this subject.

We want to complete this introduction by going briefly over the material to be covered in the rest of the present article. We shall then see how in various branches of physics collective behaviour, or quasi-particle behaviour, occurs. We shall discuss in § 2 some of the general theories developed to tackle many-body problems, in particular the one due to Tomonaga (1955 a, b). We shall not discuss a set of papers by Ecker (1955 a, b, c, d) although his discussion is germane to the subject, as this would lead us too far into the complicated problem of the applicability of the Boltzmann transport equation for the description of transport properties of systems showing collective behaviour. Ecker discusses in great detail under what conditions distribution functions can be used. He raises the point that it makes sense to use a distribution function only if it is possible to define cells in phase space so small that for all practical purposes one is dealing with functions of continuous variables, but so large that each cell contains a large number of representative points. This problem is, of course, one of the fundamental ones of statistical mechanics, but it is often too easily dismissed as being a trivial one. Ecker shows, for instance, that it may be dangerous to accept without further considerations the Debye length (see § 4) as the limiting distance for collective behaviour in plasmas. It seems to us, however, that it is probably not too dangerous to develop the various theories without taking this additional complication into account.

The main line of argument in many-body theories is the following one. One considers first of all the Hamiltonian of the system which will be expressed in, say, $3N$ Cartesian coordinates and in their conjugate momenta. One then guesses the appropriate collective coordinates, say n of them. The system now contains more coordinates $(3N+n)$ than degrees of freedom $(3N)$. This is taken care of by introducing subsidiary conditions which restrict the configuration of the system (or, in quantum mechanics, the wave functions) in such a way that the (expectation) values of the collective coordinates are equal to the (expectation) values of the appropriate functions of the original Cartesian coordinates. Often terms are added to the Hamiltonian containing the collective coordinates, which possess zero (expectation) values because of the subsidiary conditions. Instead of working explicitly with subsidiary conditions one can build them into the theory by using suitably generalized delta functions (Brenig 1956, see below). It is clearly imperative for the success of such theories that the resulting splitting of the Hamiltonian into two parts, one depending on the collective coordinates only and one depending on the (quasi-) particle coordinates only, is complete, or, if not complete, is so nearly complete that to a very good approximation interaction terms may be neglected. It may be that a sufficient condition for this to happen is the existence of an energy gap between the lowest excited collective state and the lowest excited individual state. After one has obtained the Hamiltonian the further development depends on the problem considered. In many cases the theory has not yet been fully developed along the lines indicated here.

The first example of collective behaviour we shall discuss is that of sound waves (§ 3). The sound waves themselves, when quantized, are a good example of quasi-particles, but we shall be mainly concerned with them as a manifestation of collective behaviour. We shall discuss Penrose's paper (1954) on sound waves as well as Tomonaga's considerations (1950). To describe the collective behaviour we introduce collective coordinates, which in this particular case are the Fourier components of the density fluctuations. These are particularly useful collective coordinates and we shall encounter them also in §§ 4 and 7.

In § 4 we consider briefly plasma oscillations. They were discussed before the war by Tonks and Langmuir (1929), and Steenbeck (1932) suggested that they might occur also in metals. After the war Bohm and Gross (1949 a, b) discussed the theory of plasma oscillations in gas discharges following ideas suggested earlier by Vlasov (1945 ; see also Landau 1946) and Bohm and Pines (1951, 1953, Pines and Bohm 1952, Pines 1953, 1954, 1955, 1956) extended it to metals. As far as the longitudinal oscillations are concerned, they are again described by the Fourier components of the electron density fluctuations. In the original papers the electrons were considered to be free, but recently Wolff (1953), Hubbard (1954), Mott (1954), Kanazawa (1955), and Raimes (1957) have considered the influence of the periodic lattice. As the longitudinal plasma oscillations in metals are discussed in detail by Raimes in a review in the present volume of the *Reports on Progress in Physics*, our discussion in § 4 will be mainly concerned with the plasma oscillations occurring in gas discharges.

In the theory of ferromagnetism and antiferromagnetism spin waves, first introduced by Bloch (1930), have turned out to be a powerful tool. We shall discuss them in § 5. In the ground state of a ferromagnetic all (N) spins will be parallel. The first excited state will consist of one antiparallel spin among the N–1 parallel ones. As the energy of this state will be independent of the position of the one antiparallel one, it is advantageous to use a wave picture in which the odd man out travels through the lattice. This picture is adequate as long as the number of antiparallel spins is small, since in that case we may assume the spin waves to travel independently through the lattice. The spin wave quanta, sometimes called ferromagnons, form a gas of practically independent quasi-particles. An interesting detail in this case is that the ferromagnons are bosons, while the original particles, the electrons responsible for the ferromagnetic behaviour, were fermions.

Quasi-particles similar to the ferromagnons occur in the theory of paramagnetic dielectrics. Pomerantschuk (1941 a, b, Pomerantschuk and Achieser 1944) has introduced ' magnons ' in the theory of heat conduction in these substances. A magnon is the wave-like excitation which occurs when one of the constituent magnetic moments in the solid deviates from its equilibrium value. We do not have the space to discuss these quasi-particles in more detail.

We may just mention here the excitations occurring in non-metallic substances called ' polarons ' and ' excitons '. Recently Meyer (1956 b ; see also Anselm 1951) has discussed these excitations in some detail, and also introduced a third kind, the ' excitarons ' (Meyer 1956 a). Let us consider for a moment an insulator. A substance is an insulator if its bands are fully filled up to a certain band, and if there is a forbidden gap of at least 1 eV between the last filled band and the first empty one (conduction band). In an insulator the gap is too large for an appreciable number of electrons to be present in the conduction band. However, there will be some electrons in the conduction band and at the same time some holes in the valence band. The lowest excited states of an insulator are those where the conduction electrons and the holes are correlated. Such states are called excitons. The correlated electron–hole pair can travel through the substance as a unit. It is suspected that a strong interaction between the electrons and the lattice vibrations in a metal is responsible for superconducting behaviour (Bardeen 1950 a, b, 1951 a, b, Fröhlich 1950 ; see also Salam 1954, Bonch-Bruevich 1956), the most important clue being the isotope effect. If such a strong interaction is present in a polar crystal, the low-lying stationary states of a conduction electron in such a crystal will correspond to an electron moving through the lattice accompanied by a certain amount of polarization of the lattice. This polarization can be described in terms of lattice vibrations, which means that we can describe these states as the motion of a quasi-particle which in this case is an electron plus a ' phonon cloud '. These quasi-particles were called ' polarons ' by Pekar (1946). If the electron–phonon interaction is really strong it will also influence the exciton states and a new quasi-particle, the ' excitaron ', which is an exciton surrounded by a phonon cloud will result. As the methods used in describing excitons and polarons are completely different from the methods applied in the topics treated in detail in this paper we shall not discuss excitons and polarons in detail ; we hope to discuss them elsewhere.

In §§ 4 and 5 we are concerned with solid state problems, but in §§ 6 and 7 we discuss some problems of collective behaviour in other fields of physics. It has been realized for some time now that nuclei apart from showing particle behaviour of the constituent nucleons—exhibited most clearly by the data which led Mayer (1948) to consider anew Elsasser's shell model (1933, 1934 a, b ; see also Guggenheimer 1934 a, b)—also showed collective behaviour. This collective behaviour comes to the fore when we consider those aspects of nuclei which could be understood—at least qualitatively—in terms of the liquid drop model. The collective motion of the nucleons can be described for many purposes in terms of surface oscillations, and we shall discuss these aspects of the collective behaviour in § 6. In § 7, finally, we shall discuss Feynman's theory of liquid helium (1953 a, b, 1954, 1955, Feynman and Cohen 1956) and its extension by Brenig (1956).

§ 2. GENERAL CONSIDERATIONS

We shall, to begin with, consider quite generally the problem of a system consisting of many interacting particles. After that we shall specialize and discuss slightly more specific problems. In so far as the essential features are shown in a classical context we shall use the classical language, but for typical quantum effects we shall have recourse to wave mechanics.

The general problem we are concerned with is the solution of the equations of motion of a system of N interacting particles. These particles may all be of the same kind, or there may be several kinds of particles. Let us for the moment assume that they are all of the same kind. For the sake of simplicity we shall also assume that we may use for the potential energy the sum of the potential energies of pairs of particles. In terms of Cartesian coordinates, the Hamiltonian H of the system is given by the equation

$$H = T+U, \quad T = \tfrac{1}{2}\sum_i m_i \dot{\mathbf{x}}_i^2, \quad U = \sum_{i<j} U_{ij}(\mathbf{x}_i, \mathbf{x}_j), \qquad \ldots\ldots(2.1)$$

where T and U are the kinetic and potential energy, $\mathbf{x}_i$ the coordinates of the ith particle and m_i its mass. The equations of motion are of the form

$$m_i \ddot{\mathbf{x}}_i = -\nabla_i U. \qquad \ldots\ldots(2.2)$$

For many properties of the system it is sufficient, if we can find the extension in phase, Ω, to use Gibbs' terminology (1902 ; see ter Haar 1954 a, p. 101), as a function of energy, because we can then evaluate—at least in principle—the partition function Z of the system which is given by the equation

$$Z = \int e^{-\beta\epsilon}\, d\Omega = \int e^{-\beta\epsilon} \frac{\partial \Omega}{\partial \epsilon}\, d\epsilon, \qquad \ldots\ldots(2.3)$$

where $d\Omega$ is an element of extension in phase, and β denotes here and henceforth $1/\boldsymbol{k}T$ where $\boldsymbol{k}$ is Boltzmann's constant, and T is the absolute temperature. All equilibrium properties of the system can be calculated, once Z is known as a function of β (or T) and of the external parameters which are contained in the potential energy.

In the case of a system where quantum effects are of importance, we have instead of equations (2.2)

$$H_{\text{op}}\Psi = E\Psi, \quad H_{\text{op}} = \sum_i T_{i\text{op}} + \sum_{i<j} U_{ij}, \qquad \ldots\ldots(2.4)$$

where the kinetic energy operators $T_{i_{op}}$ are obtained from the T_i in the usual way, and where Ψ is the wave function of the whole system and E the energy. The problem now is to find the energy levels E_k, or for most cases the energy level density $\rho(E)$, since now Z is given by the equation

$$Z = \sum_k g_k \exp(-\beta E_k) \simeq \int e^{-\beta E} \rho(E)\, dE, \qquad \ldots\ldots(2.5)$$

where g_k is the degree of degeneracy, or weight, of the level E_k.

Equations (2.2) and (2.4) are clearly too complicated in general to be solved directly, except in the trivial case where $U = 0$. We must thus have recourse to other methods to find a solution. The first one is the Hartree–Fock method (see, for instance, Schiff 1949, chap. XI) which works essentially as follows. We know that the various particles are interacting and that thus the wave function will be a complicated expression in the $\mathbf{x}_i$. However, if the interaction is not too strong, or rather, if close encounters can to a first approximation be neglected, we can approximate the situation by assuming the various particles to move in a potential field which is produced by all the other particles. The total wave function is then a product of single particle wave functions, or rather, a suitably symmetrized or antisymmetrized function built up out of products of single particle wave functions. It will be seen immediately that the background potential will depend on the ' states ' occupied by the various particles, and it must thus be determined from a consistency relation. Let $\phi_l(i)$ indicate the normalized wave function of the ith particle in the lth state. This means that Ψ is assumed to be written in the form

$$\Psi = \Pi_i \phi_{l_i}(i). \qquad \ldots\ldots(2.6)$$

The $\phi_l(i)$ satisfy the equation

$$(E_l - T_i - V_i)\phi_l(i) = 0, \qquad \ldots\ldots(2.7)$$

and this set of equations will be equivalent to equation (2.4) provided V_i, the pseudo potential, is given by the equation

$$V_i = \sum_{j(\neq i)} \int \phi_{l_j}(j)\, U_{ij}\, \phi_{l_j}(j)\, d\mathbf{x}_j, \qquad \ldots\ldots(2.8)$$

as can be seen by inspection, where here and henceforth $d\mathbf{x}$ indicates a volume in coordinate space. We note from equation (2.8) that, indeed, V_i depends on the states in which the other particles are. The way to solve the problem is now to assume an expression for V_i, then to solve equation (2.7) and to check whether the chosen V_i will lead to wave functions $\phi_l(i)$ which, substituted into equation (2.8), using the known expression for U_{ij}, will lead to V_i which are the same as the original V_i. If this is not the case, one can use various approximation methods to improve the situation. We do not wish to pursue this line any further. Gombas and collaborators have in recent years used the Hartree–Fock method for atoms and solids (Gombas 1954, 1956 a, b, Gombas and Szépfalusy 1955, Gombas and Ladányi 1955, Gombas and Kunvári 1955 and other articles referred to in these, especially Gombas 1949) and Slater and collaborators have used it in solid state research (mainly in unpublished M.I.T. reports ; see also Slater 1951 a, b, 1953). As we mentioned in § 1, Brueckner has shown a way to modify the Hartree–Fock method so as to make it applicable to the case of the strongly interacting nucleons (for an account of Brueckner's work see Eden 1956, Bethe 1956). Similar methods have been developed by Riesenfeld and Watson (1956) and we may also refer to papers by Jastrow (1955) and by Kinoshita and Nambu (1954). All these methods try to express the behaviour of the system in terms of the original coordinates. In classical terms, the situation is described in terms of single particle orbits, and the collective aspects enter through the smoothed-out potential. The real collective modes and the quasi-particles do not appear clearly in this approach.

A different way of approach is the one developed by Tomonaga (1955 a, b ; see also Yevick and Percus 1956, Percus and Yevick 1956, Skinner 1956, Brenig 1956). In this method one assumes that one knows the form the collective modes will take and then proceeds to introduce suitable coordinates pertaining to those modes. Let the (infinitesimal) displacement of the particles $\delta\mathbf{x}_i$ be given by the equations

$$\delta\mathbf{x}_i = \epsilon\nabla_i\phi(\mathbf{x}_i), \qquad \ldots\ldots(2.9)$$

where $\phi(\mathbf{x})$ is a function which will determine the form of the collective motion, and where ϵ is an (infinitesimal) constant.

The first step is now to find the momentum π which is conjugate to the collective coordinate corresponding to the motion (2.9). This momentum can be found from the requirement that the change δf of an arbitrary function f of the $\mathbf{x}_i$ is given by the equation

$$\{\pi, f\} = \delta f\left[= \sum_i (\delta\mathbf{x}_i \cdot \nabla_i f)\right], \qquad \ldots\ldots(2.10)$$

where $\{a, b\}$ is the Poisson bracket of a and b,

$$\{a, b\} = \sum_i \left(\frac{\partial a}{\partial q_i}\frac{\partial b}{\partial p_i} - \frac{\partial a}{\partial p_i}\frac{\partial b}{\partial q_i}\right), \qquad \ldots\ldots(2.11)$$

if the p_i are the generalized momenta conjugate to the generalized coordinates q_i. It is easily verified that equation (2.10) is satisfied, if we take

$$\pi = \sum_i (\mathbf{p}_i \cdot \nabla_i\phi(\mathbf{x}_i)), \qquad \ldots\ldots(2.12)$$

where $\mathbf{p}_i = m_i\mathbf{x}_i$ is the linear momentum of the ith particle.

The next step is more complicated. We must now find the collective coordinate ξ. The first requirement for ξ is to satisfy the equation

$$\{\pi, \xi\} = -1, \qquad \ldots\ldots(2.13)$$

which is the condition for π and ξ to be canonically conjugate. However, there is another requirement which follows from the physics of the problem. The behaviour of the system will be described by other coordinates—let us call them for the moment particle coordinates—apart from ξ. Let η be such a coordinate. We must then require first of all that η satisfies the equation

$$\{\pi, \eta\} = 0, \qquad \ldots\ldots(2.14)$$

but also that the surfaces η = constant are orthogonal to the surfaces ξ = constant as otherwise the kinetic energy will contain cross terms which would lead to a complicated entanglement of particle and collective coordinates, making it impossible to separate them. This orthogonality means that

$$\sum_i (\nabla_i \xi . \nabla_i \eta) = 0, \qquad \ldots\ldots(2.15)$$

while equation (2.14) can be written in the form

$$\sum_i (\nabla_i \phi . \nabla_i \eta) = 0. \qquad \ldots\ldots(2.16)$$

There are now two ways of finding ξ. The first one is that suggested by Tomonaga, namely, to take for ξ the expression

$$\xi = A^{-1} \sum_i \phi(\mathbf{x}_i). \qquad \ldots\ldots(2.17)$$

In that case equation (2.15) is satisfied because it is identical with equation (2.16). We must now see whether equation (2.13) is satisfied, and, if not, whether or not it is badly violated. The Poisson bracket $\{\pi, \xi\}$ is easily evaluated, and we have

$$\{\pi, \xi\} = -A^{-1} \sum_i (\nabla_i \phi . \nabla_i \phi). \qquad \ldots\ldots(2.18)$$

We now take

$$A = \left\langle \sum_i (\nabla_i \phi . \nabla_i \phi) \right\rangle, \qquad \ldots\ldots(2.19)$$

where the brackets $\langle\,\rangle$ indicate an average over all possible configurations in coordinate space. For most functions ϕ we may expect the dispersion of the expression $\sum_i (\nabla_i \phi . \nabla_i \phi)$ to be negligibly small, that is, we may expect

$$\left[\left\langle \left\{ \sum_i (\nabla_i \phi . \nabla_i \phi) \right\}^2 \right\rangle - \left\langle \sum_i (\nabla_i \phi . \nabla_i \phi) \right\rangle^2 \right] \Big/ \left\langle \sum_i (\nabla_i \phi . \nabla_i \phi) \right\rangle^2$$

to be of the order of magnitude N^{-1}, where N is the number of particles (Tomonaga (1955 a) gives a proof of this statement). This means that equation (2.13) is satisfied as long as we may neglect terms of the order $N^{-\frac{1}{2}}$.

The second method of finding ξ is the one used by Brenig (1956). In this case one chooses ξ so as to satisfy equation (2.13) exactly, but one satisfies equation (2.15) only up to terms of relative order $N^{-\frac{1}{2}}$. We refer to § 7 for details of this procedure.

We have now constructed the collective coordinates and momenta by using equations (2.12) and (2.17). The next step is to express the Hamiltonian in terms of π, ξ, and the individual, or internal, coordinates. We consider first of all the kinetic energy and we try to find a constant I such that we can write

$$T = \sum_i (\mathbf{p}_i^2/2m_i) = T_{\text{in}} + T_{\text{c}}, \quad T_{\text{c}} = \pi^2/2I, \qquad \ldots\ldots(2.20)$$

where T_{in} does not contain π so that

$$\{T_{\text{in}}, \xi\} = 0. \qquad \ldots\ldots(2.21)$$

From equations (2.20), (2.21), (2.17) and (2.12) it follows that I is given by the equation

$$I = mA, \qquad \ldots\ldots(2.22)$$

where we have put all m_i equal to m. We now have for T_{in} the equation

$$T_{\text{in}} = \sum_i (\mathbf{p}_i^2/2m_i) - \left[\sum_i (\mathbf{p}_i . \nabla_i \phi) \right]^2 \Big/ 2mA. \qquad \ldots\ldots(2.23)$$

Although $T_{\rm in}$ does not contain π it will in general still contain ξ and we should like to know in what way it depends on ξ. In many cases it will take a finite amount of energy to excite the collective modes which means that their amplitude will be small. In such cases we can expand $T_{\rm in}$ in a power series in ξ and only retain terms up to ξ^2. Unfortunately, if we do not want to introduce the internal coordinates explicitly, we only know $T_{\rm in}(\xi), T_{\rm in}'(\xi), T_{\rm in}''(\xi),\ldots$ (primes indicate differentiation with respect to ξ), and not $T_{\rm in}(0), T_{\rm in}'(0), T_{\rm in}''(0),\ldots$ which enter into a Taylor expansion for $T_{\rm in}(\xi)$. Instead, we use the identity

$$T_{\rm in}(\xi) = \{T_{\rm in}(\xi) - \xi T_{\rm in}'(\xi) + \tfrac{1}{2}\xi^2 T_{\rm in}''(\xi)\} + \xi\{T_{\rm in}'(\xi) - \xi T_{\rm in}''(\xi)\} + \tfrac{1}{2}\xi^2\{T_{\rm in}''(\xi)\}, \qquad \ldots\ldots(2.24)$$

where the expressions within braces are independent of ξ up to terms of the order $\xi^2(\xi^1, \xi^0)$. We can now express the braces in terms of the original coordinates $\mathbf{x}_i$ by writing

$$T_{\rm in}'(\xi) = \{\pi, T_{\rm in}\}, \quad T_{\rm in}''(\xi) = \{\pi, \{\pi, T_{\rm in}\}\}, \qquad \ldots\ldots(2.25)$$

and using expressions (2.12) and (2.17) for π and ξ.

A similar procedure is adopted to express the potential energy U as a function of ξ. We get finally

$$H = (T_{\rm in}^{(0)} + U^{(0)}) + (T_{\rm in}^{(1)} + U^{(1)})\xi + (T_{\rm in}^{(2)} + U^{(2)})\xi^2 + \pi^2/2I. \qquad \ldots\ldots(2.26)$$

The first bracket refers to the individual motion, the last bracket and the $\pi^2/2I$ term refer to the collective motion, while the second bracket refers to the interaction of the individual with the collective modes. It is clearly important for the success of the theory that this interaction term is small. In later sections we shall see several illustrations of Tomonaga's theory so that we shall only briefly show how it works in the simple case where we are dealing with the separation of the centre of mass motion. If we consider the x-component of this motion, we have for ϕ the equation

$$\phi(\mathbf{x}) = x, \qquad \ldots\ldots(2.27)$$

and the collective motion described by equation (2.9) is a simple translation. Equation (2.12) leads to

$$\pi = \sum_i p_{xi}, \qquad \ldots\ldots(2.28)$$

while ξ is given by equations (2.17) and (2.19),

$$\xi = A^{-1}\Sigma x_i, \quad A = N, \qquad \ldots\ldots(2.29)$$

that is, π is the x-component of the total linear momentum and ξ the x-coordinate of the centre of gravity (we are dealing with N particles of equal mass). Equation (2.13) is satisfied exactly, and the quantity I given by equation (2.22) is just the total mass of the system. In this particular case, H will not depend on ξ so that the second and third bracket in equation (2.26) will not occur.

Tomonaga discusses the equivalence of his method and the method of auxiliary fields used by Bohm and Pines, and by Kinoshita and Nambu. We do not wish to discuss the use of auxiliary fields and we refer therefore the reader to Tomonaga's papers for such a discussion. Zubarev (1953) and Galasiewicz (1955, 1956 a, b) also use the method of redundant variables.†

---

† After this section was written, Dr. Skyrme kindly sent me a preprint of a paper (Skyrme 1957) which discusses collective motion from a slightly different point of view. We hope to discuss his method elsewhere.

§ 3. SOUND WAVES

Consider a gas of density ρ at a pressure p with a ratio of the specific heat denoted by γ. It is well known (see, for instance, Morse 1936) that the density in such a medium is governed by the equation

$$\nabla^2 \rho - \frac{1}{c^2}\frac{\partial^2 \rho}{\partial t^2} = 0, \qquad \ldots\ldots(3.1)$$

where c is the velocity of sound given by the equation

$$c^2 \rho_0 = \gamma p_0, \qquad \ldots\ldots(3.2)$$

with ρ_0 and p_0 the equilibrium values of ρ and p. Equation (3.1) describes the propagation of sound waves, which are clearly a collective phenomenon. In order to find the most suitable coordinates to describe phonons we take the Fourier components $\rho_{\mathbf{k}}$ of the density,

$$\rho_{\mathbf{k}} = \int \rho(\mathbf{x}) \exp\,[i(\mathbf{k}\,.\,\mathbf{x})]\, d\mathbf{x}, \qquad \ldots\ldots(3.3)$$

where $\mathbf{k}$ is the wave vector. Substituting from equation (3.3) into equation (3.2) we get

$$\ddot{\rho}_{\mathbf{k}} + k^2 c^2 \rho_{\mathbf{k}} = 0, \qquad \ldots\ldots(3.4)$$

which shows that the $\rho_{\mathbf{k}}$ behave as harmonic oscillators of frequency kc (k is the absolute magnitude of $\mathbf{k}$). We shall call equation (3.4) the basic phonon equation. We may refer to a paper by Penrose (1954) for a discussion of the quantization of the phonon equation.

Equation (3.1) is a phenomenological equation and up to this point we have not introduced any assumptions as to the kind of interparticle forces which in many-body problems might lead to equation (3.1), or to equation (3.4). We shall consider a few cases and refer also to § 4 and § 7 where other cases of longitudinal waves will be discussed. We shall first of all consider the problem in general terms before going over to specific applications. We can use the general framework of § 2, and we shall consider the collective coordinates corresponding to a given wave vector $\mathbf{k}$. As we mentioned a moment ago, these will be the Fourier components of the density. As we are now dealing with a system of particles, and not with a continuum, we have for the density the equation

$$\rho(\mathbf{x}) = m \sum_j \delta(\mathbf{x} - \mathbf{x}_j), \qquad \ldots\ldots(3.5)$$

where $\delta(\mathbf{x})$ is the three-dimensional delta-function, and m the mass of the particles. From equation (3.5) it follows that

$$\rho_{\mathbf{k}} = m \sum_j \exp\,[i(\mathbf{k}\,.\,\mathbf{x}_j)], \quad \rho_{-\mathbf{k}} = m \sum_j \exp\,[-i(\mathbf{k}\,.\,\mathbf{x}_j)]. \qquad \ldots\ldots(3.6)$$

To put these coordinates on a par with the ξ's discussed in § 2 we shall denote them by ξ^+, ξ^- and normalize them slightly differently. We take the following generating functions (it must be noted that they are slightly different from the $\phi^\pm$ introduced by Tomonaga)

$$\phi^\pm = k^{-1} \exp\,[\pm i(\mathbf{k}\,.\,\mathbf{x})]. \qquad \ldots\ldots(3.7)$$

We are now dealing with complex quantities and we cannot consider ϕ^+ and ϕ^- separately, but we must consider them at the same time. We shall continue to use equation (2.17) for $\xi^{\pm}$ and we have thus

$$\xi^{\pm} = (kN)^{-1} \sum_j \exp\left[\pm i(\mathbf{k}.\mathbf{x}_j)\right], \qquad \ldots\ldots(3.8)$$

where we have put $A = N$. We shall show presently that for this particular choice of A, equation (2.13) will be satisfied.

Instead of equation (2.12) we now use for $\pi^{\pm}$ the equations

$$\pi^{\pm} = \sum_j (\mathbf{p}_j . \nabla_j \phi^{\mp}(\mathbf{x}_j)), \qquad \ldots\ldots(3.9)$$

which leads to the expressions

$$\pi^{\pm} = \mp (i/k) \sum_j (\mathbf{p}_j . \mathbf{k}) \exp\left[\mp i(\mathbf{k}.\mathbf{x}_j)\right]. \qquad \ldots\ldots(3.10)$$

It follows easily that

$$\{\pi^+, \xi^+\} = \{\pi^-, \xi^-\} = -1, \quad \{\xi^+, \xi^-\} = 0, \qquad \ldots\ldots(3.11)$$

while

$$\{\pi^+, \pi^-\} = -2i \sum_j (\mathbf{p}_j . \mathbf{k}), \quad \{\pi^+, \xi^-\} = -N^{-1} \sum_j \exp\left[-2i(\mathbf{k}.\mathbf{x}_j)\right] = \{\pi^-, \xi^+\}^*, \qquad \ldots\ldots(3.12)$$

where * indicates the conjugate complex. We notice that if we are working in the centre-of-mass system so that $\sum_j \mathbf{p}_j = 0$, $\{\pi^+, \pi^-\} = 0$. Secondly, although $\{\pi^+, \xi^-\}$ and $\{\pi^-, \xi^+\}$ are not vanishing, they are vanishing in the mean. This is meant as follows. The Poisson bracket $\{\pi^+, \xi^-\}$ is of an order $N^{-\frac{1}{2}}$ smaller than $\{\pi^+, \xi^+\}$, and to a first approximation can thus be neglected. Under those assumptions all Poisson brackets of equations (3.12) vanish, and the $\pi^+, \pi^-, \xi^+, \xi^-$ form a suitable set of canonical coordinates.

The next step is to express the Hamiltonian in terms of the $\xi^{\pm}$ and $\pi^{\pm}$. As these coordinates are complex, we use for the collective kinetic energy the expression

$$T_c = \pi^+ \pi^- / Nm, \qquad \ldots\ldots(3.13)$$

which on inspection shows to satisfy equation (2.21). It follows that the T_{in} is given by the equation

$$T_{\text{in}} = \sum_j (p_j^2/2m) - (2mNk^2)^{-1} \left\{ \sum_j (\mathbf{p}_j . \mathbf{k}) \exp\left[-i(\mathbf{k}.\mathbf{x}_j)\right] \right\} \left\{ \sum_j (\mathbf{p}_j . \mathbf{k}) \exp\left[i(\mathbf{k}.\mathbf{x}_j)\right] \right\}. \qquad \ldots\ldots(3.14)$$

For the potential energy we shall again assume additivity and central forces so that we have as in equation (2.1)

$$U = \tfrac{1}{2} \sum_{i,j} U_{ij}(r_{ij}) = \tfrac{1}{2} \sum_{i,j} U(r_{ij}). \qquad \ldots\ldots(3.15)$$

Furthermore we shall use the Fourier expansion of $U(r_{ij})$,

$$U(r_{ij}) = \sum_{\mathbf{k}} f(k) \exp\left[i(\mathbf{k}.\mathbf{x}_{ij})\right], \qquad \ldots\ldots(3.16)$$

where $f(k)$ depends on the absolute magnitude k of $\mathbf{k}$ only.

Using the method described in § 2 we can express the energy in terms of $\xi^{\pm}, \pi^{\pm}$, and an individual part. Neglecting the ξ-dependent terms in the kinetic energy the Hamiltonian is given by the equation

$$H = H_{\text{in}} + (\pi^+\pi^-/Nm) + N^2k^2f(k)\,\xi^+\xi^-, \qquad \ldots\ldots(3.17)$$

corresponding to a frequency ω of the collective modes given by the equation

$$\omega^2 = Nk^2f(k)/m. \qquad \ldots\ldots(3.18)$$

We refer to Tomonaga for a derivation of equation (3.17)—which is equivalent to equation (3.4)—and for a discussion of the assumptions introduced in the derivation.

There are two simple cases—both one-dimensional—which we shall discuss in this section. The first one is that of the linear crystal, and the second one is that of a nearly completely degenerate perfect one-dimensional Fermi gas (Tomonaga 1950; see also Bloch 1934).

In the first case we use the Born–von Kármán model (1912, 1913) of a crystal. Let the crystal be in the x-direction, and we assume the equilibrium state to be one where the atoms are equally spaced. Let L be the repeating distance of the usual periodic boundary conditions along which there are N atoms. We number the particles in their order along the crystal and assume all $x_{j+1}-x_j-(L/N)$ to be small. Assuming harmonic forces—or neglecting third and higher order terms in a Taylor expansion—we get for the Hamiltonian the expression

$$H = \sum_j (p_j^2/2m) + \tfrac{1}{2}a\sum_j [x_{j+1}-x_j-(L/N)]^2, \qquad \ldots\ldots(3.19)$$

where j runs from 1 to N and where $x_{N+1} = x_1$ in accordance with the periodicity requirement. We shall make the additional assumption that not only the quantities in square brackets in equation (3.19) are small, but that also the $x_{j+1}-jL/N$ are small. In that case we have for the $\xi^\pm$ from equation (3.8)

$$\begin{aligned}\xi^\pm &= (kN)^{-1}\sum_j \exp(\pm ikx_j) = (kN)^{-1}\sum_j \exp(\pm ikjL/N)\exp[\pm ik(x_j-jL/N)]\\ &\simeq \text{constant} \pm (i/N)\sum_j [x_j-(jL/N)]\exp(\pm ikjL/N). \qquad \ldots\ldots(3.20)\end{aligned}$$

To the same approximation the $\pi^\pm$ are given by the equations [see equation (3.10)]

$$\pi^\pm = \mp i\sum_j p_j \exp(\mp ikjL/N). \qquad \ldots\ldots(3.21)$$

We drop the constant term of the $\xi^\pm$. We know that we must restrict k to the N values (we assume N to be odd for the sake of simplicity)

$$k = 2l\pi/L, \quad l = 0, \pm 1, \pm 2, \ldots, \quad \pm\tfrac{1}{2}(N-1). \qquad \ldots\ldots(3.22)$$

These values of k follow from the periodicity condition. If $\xi_{\pm k}$ indicate $\xi^\pm$ for a given wave number k, we can use equation (3.20) to write

$$x_j-(jL/N) = \mp i\sum_k \xi_{\pm k}\exp(\mp ikjL/N), \qquad \ldots\ldots(3.23)$$

and similarly

$$p_j = \pm(i/N)\sum_k \pi_{\pm k}\exp(\pm ikjL/N). \qquad \ldots\ldots(3.24)$$

Finally, we have

$$\begin{aligned}x_j-x_{j-1}-L/N &= \mp i\sum_k \xi_{\pm k}[1-\exp(-ikL/N)]\exp(\mp ikjL/N)\\ &\simeq \pm\sum_k (kL/N)\,\xi_{\pm k}\exp(\mp ikjL/N). \qquad \ldots\ldots(3.25)\end{aligned}$$

Using equations (3.24), (3.25) and (3.19) we get for the Hamiltonian the expression

$$H = \tfrac{1}{2}\sum_k [(\pi_{+k}\pi_{-k}/Nm) + (ak^2L^2/N)\,\xi_{+k}\xi_{-k}], \qquad \ldots\ldots(3.26)$$

and we see that for each wave number k we have a term in the Hamiltonian corresponding to equation (3.4). The frequency ω is given by the equation

$$\omega^2 = aL^2k^2/mN^2, \qquad \ldots\ldots(3.27)$$

and the sound velocity c by the equation

$$c^2 = aL^2/mN^2. \qquad \ldots\ldots(3.28)$$

We notice from equations (3.27) and (3.28) that our approximations correspond to the Debye approximation (1912) $\omega = c\,|k|$.

Let us now consider a degenerate one-dimensional perfect gas of fermions. In this case we must first of all evaluate the ξ-dependent terms in the kinetic energy, and secondly use quantum rather than classical mechanics.

To get the ξ-dependent terms in T we use equation (3.24), but as we are now dealing with complex ξ we must expand T in terms of both ξ^+ and ξ^-. We can use equation (3.13) for T_c and get from equation (3.14) in the one-dimensional case for T_{in} the equation

$$T_{\text{in}} = \sum_j (p_j^2/2m) - (2mN)^{-1}\Big[\sum_j p_j \exp(-ikx_j)\Big]\Big[\sum_j p_j \exp(ikx_j)\Big]. \qquad \ldots\ldots(3.29)$$

The Poisson brackets which occur in equation (2.25) are easily evaluated and we have

$$\begin{aligned} \{\pi^\pm, T_{\text{in}}\} &= -(ik/m)\sum_j p_j^2 \exp(\mp ikx_j)\,; \\ \{\pi^+,\{\pi^+, T_{\text{in}}\}\} = \{\pi^-,\{\pi^-, T_{\text{in}}\}\}^* &= (k^2/m)\sum_j p_j^2 \exp(-2ikx_j)\,; \end{aligned} \qquad \ldots\ldots(3.30)$$

$$\{\pi^+,\{\pi^-, T_{\text{in}}\}\} = (3k^2/m)\sum_j p_j^2. \qquad \ldots\ldots(3.31)$$

In deriving the first of equations (3.30) we have assumed that $\sum_j \mathbf{p}_j = 0$, that is, that we are working in the centre-of-mass system.

The transition to quantum mechanics is simple, and should really have been made earlier. All p_j are treated as operators $-i\hbar\partial/\partial x_j$, and all Poisson brackets are replaced by commutators. One must be careful in defining π because it contains non-commuting quantities. However, if we are dealing with long wavelength modes so that $\hbar k$ is small compared to the average value of p_j, we get essentially equations (3.30) and (3.31) (see Tomonaga 1955 b).

We must now consider the final expression for T_{in} as given by the equivalent of equation (2.24). We see that $\{\pi^\pm, T_{\text{in}}\}$ will be multiplied by $\xi^\pm$, $\{\pi^\pm,\{\pi^\pm, T_{\text{in}}\}\}$ by $\xi^\pm\xi^\pm$, and $\{\pi^\pm,\{\pi^\mp, T_{\text{in}}\}\}$ by $\xi^+\xi^-$. It is reasonable (and usual) to assume that any sum over j of terms containing factors $\exp(\pm ikx_j)$ will be smaller than the same sum without these factors by a factor of the order of $N^{\frac{1}{2}}$. (In the case of degenerate Fermi gas the factor is even of the order N.) This means that we can neglect in our final expression for T_{in} $\{\pi^+,\{\pi^+, T_{\text{in}}\}\}$ and $\{\pi^-,\{\pi^-, T_{\text{in}}\}\}$. We thus get for T_{in}

$$T_{\text{in}} = T_0 + T_1^{+}\xi^+ + T_1^{-}\xi^- + T_2^{++}\xi^+\xi^+ + T_2^{+-}\xi^+\xi^- + T_2^{--}\xi^-\xi^-, \qquad \ldots\ldots(3.32)$$

with

$$\left.\begin{aligned} T_2^{++} &= T_2^{--} = 0, \quad T_2^{+-} = (3k^2/m)\sum_j p_j^2, \\ T_1^{+} = T_1^{-*} &= -(k/m)\sum_j p_j^2 \exp(-ikx_j) \\ &\quad -(3k/mN)\left[\sum_j \exp(-ikx_j)\right]\left[\sum_j p_j^2\right], \\ T_0 &= \sum_j (p_j^2/2m) - (2mN)^{-1}\left|\sum_j p_j \exp(-ikx_j)\right|^2 \\ &\quad +(mN)^{-1}\left\{\left[\sum_j p_j^2 \exp(-ikx_j)\right]\left[\sum_j \exp(ikx_j)\right]\right. \\ &\quad \left. +\left[\sum_j p_j^2 \exp(ikx_j)\right]\left[\sum_j \exp(-ikx_j)\right]\right\} \\ &\quad +(3/mN^2)\left(\sum_j p_j^2\right)\left|\sum_j \exp(ikx_j)\right|^2. \end{aligned}\right\} \qquad \ldots\ldots(3.33)$$

We must now introduce the assumption of a nearly degenerate Fermi gas, in order to evaluate the sums in equation (3.33). Let p_0 be the momentum corresponding to the Fermi level. In that case we have

$$p_0 = hN/L, \qquad \ldots\ldots(3.34)$$

where h is Planck's constant, L the volume (length) of the system, and N the number of particles in the system. The sum occurring in the expression for T_2^{+-} is easily evaluated in the case of degeneracy, as it is then $2m$ times the sum of all energies up to the Fermi level. We get in this way for T_2^{+-} the equation

$$T_2^{+-} = Nk^2p_0^2/m = h^2k^2N^3/mL^2. \qquad \ldots\ldots(3.35)$$

We must now evaluate the expressions for $T_1^{\pm}$. The second term is evaluated in the same way as T_2^{+-}, and to evaluate the first term, or rather its expectation value, we note that the operation of $\exp(ikx_j)$ corresponds to the transition from a state of momentum p_j to a state of momentum $p_j + \hbar k$. This means that p_j must be below or near p_0 and $p_j + \hbar k$ near or above p_0. We shall assume $\hbar k$ to be small compared with p_0. This means that in the sums $\sum_j p_j^2 \exp(\pm ikx_j)$ we may replace p_j by p_0 and we get thus for $T_1^{\pm}$ the equations

$$T_1^{\pm} = -(k/m)\left[\sum_j \exp(\mp ikx_j)\right]\left\{p_0^2 - (3/N)\sum_j p_j^2\right\} = 0. \qquad \ldots\ldots(3.36)$$

If $U = 0$, we have for the total Hamiltonian the expression

$$H = T_0 + (h^2k^2N^3/mL^2)\,\xi^+\xi^- + (\pi^+\pi^-/Nm), \qquad \ldots\ldots(3.37)$$

corresponding to a frequency

$$\omega = hkN/mL. \qquad \ldots\ldots(3.38)$$

In the three-dimensional case $T_1^{\pm}$ is not equal to zero and there is a strong interaction between the individual and the collective modes. If there are interactions we must consider the ξ-dependent terms in U as well. In the case where $U_2^{+-}\xi^+\xi^-$ and $T_2^{+-}\xi^+\xi^-$ are the most important collective terms we find by combining equations (3.37) and (3.17) that the frequency is given by the equation

$$\omega^2 = (Nk^2f(k)/m) + (h^2k^2N^2/m^2L^2). \qquad \ldots\ldots(3.39)$$

§ 4. PLASMA OSCILLATIONS

Raimes (1957) discusses in detail in the same volume as the present article plasma oscillations in metals, and there is clearly no point in going over the same ground here. However, we want to show how the longitudinal plasma oscillations fit in the general framework of Tomonaga's theory.

We have already derived most formulae which we need, and we only need generalize equation (3.31) for $\{\pi^+, \{\pi^-, T_{\text{in}}\}\}$ to the case of a three-dimensional system. The result is

$$\{\pi^+, \{\pi^-, T_{\text{in}}\}\} = \frac{3}{m}\sum_j (\mathbf{p}_j^2 . \mathbf{k})^2 = 3m \sum_j (\mathbf{v}_j . \mathbf{k})^2 = 3Nm\langle(\mathbf{v}_j . \mathbf{k})^2\rangle, \qquad \ldots\ldots(4.1)$$

where now and throughout § 4 $\langle\rangle$ will indicate an average over all the particles.

Neglecting the terms in ξ^+ and ξ^- (this corresponds to the random phase approximation of Pines and Bohm 1952), and combining equations (3.17) with the $\xi^+\xi^-$ term of the kinetic energy, we get for the collective part, H_c, of the Hamiltonian the expression

$$H_c = (\pi^+\pi^-/Nm) + \{N^2k^2f(k) + 3Nm\langle(\mathbf{v}_j . \mathbf{k})^2\rangle\}\xi^+\xi^-, \qquad \ldots\ldots(4.2)$$

corresponding to a frequency given by the equation

$$\omega^2 = (Nk^2f(k)/m) + 3\langle(\mathbf{v}_j . \mathbf{k})^2\rangle. \qquad \ldots\ldots(4.3)$$

In the case of Coulomb forces we have

$$U_{ij} = -e^2/r_{ij}, \quad f(k) = 4\pi e^2/k^2, \qquad \ldots\ldots(4.4)$$

and we get from equation (4.3) the well-known dispersion relation (Bohm and Gross 1949 a, Pines and Bohm 1952 ; see also Thomson and Thomson 1933, van Kampen 1955, Clemmow and Willson 1956)

$$\omega^2 = \omega_{\text{p}}^2 + 3\langle(\mathbf{v}_j . \mathbf{k})^2\rangle, \qquad \ldots\ldots(4.5)$$

where ω_{p} is the well-known plasma frequency (Tonks and Langmuir 1929)

$$\omega_{\text{p}}^2 = 4\pi Ne^2/m. \qquad \ldots\ldots(4.6)$$

We see that the frequency ω will be approximately equal to ω_{p} as long as

$$4\pi Ne^2/m \gg 3\langle(\mathbf{v}_j . \mathbf{k})^2\rangle \simeq k^2\boldsymbol{k}T/m, \qquad \ldots\ldots(4.7)$$

where $\boldsymbol{k}$ is the Boltzmann constant and T the temperature of the system. Inequality (4.7) also gives the condition for collective behaviour to occur (see, for instance, Pines 1955). From inequality (4.7) it follows that collective behaviour occurs for phenomena of wavelength longer than the Debye length (Debye and Hückel 1923) λ_{D} given by the equation

$$\lambda_{\text{D}}^2 = \boldsymbol{k}T/4\pi Ne^2. \qquad \ldots\ldots(4.8)$$

We must emphasize that the equations must be suitably changed if we are dealing with the electrons in a metal which forms a nearly degenerate Fermi gas (Bohm and Pines 1953, Pines 1955, Raimes 1957). We may mention that Groschwitz and Siebertz (1956) have discussed the occurrence of plasma oscillations in semiconductors.

There is one further point we wish to make here. If there are no collective modes, the potential energy is given by the expression (see equations (3.15) and (3.16))

$$U = \tfrac{1}{2} \sum_{i,j} \sum_{\mathbf{k}} f(k) \exp [i(\mathbf{k} \cdot \mathbf{x}_{ij})]. \qquad \ldots\ldots(4.9)$$

However, if there are longitudinal collective waves corresponding to $\xi^{\pm}$ with wave vectors $\mathbf{k}$ of absolute magnitude up to k_c we have instead of U $U^{(0)}$ (see equation (2.26)) as the potential energy of the individual modes. Tomonaga (1955 b) has shown that $U^{(0)}$ is given by the expression

$$U^{(0)} = \tfrac{1}{2} \sum_{i,j} U_{ij}{}^{(0)}, \; U_{ij}{}^{(0)} = \sum_{|\mathbf{k}| > k_c} f(k) \exp [i(\mathbf{k} \cdot \mathbf{x}_{ij})]. \qquad \ldots\ldots(4.10)$$

We see here the 'screening' effect: all terms corresponding to long-range interactions (long wavelengths) have dropped out.

§ 5. SPIN WAVES

Ferromagnetism is a very complicated and fascinating physical phenomenon. In his review of the subject Stoner (1948) stated: "The rich diversity of ferromagnetic phenomena, the perennial challenge to skill in experiment and to physical insight in coordinating the results, the vast range of actual and possible applications of ferromagnetic materials, and the fundamental character of the essential theoretical problems raised have all combined to give ferromagnetism a width of interest which contrasts strongly with the apparent narrowness of its subject matter, namely, certain particular properties of a very limited number of substances." From the theoretical point of view one has a fair idea where to look for the reason for ferromagnetic behaviour. The reason is essentially an exchange interaction proposed independently by Frenkel and by Heisenberg (see also Bloch 1929). Such an exchange interaction will give rise to the so-called Weiss molecular field (for a more or less elementary account of ferromagnetism see, for instance, Kittel 1953, chap. 10). One of the simplest models is the one proposed by Ising (1925) which can also be applied to order–disorder (see, for instance, Newell and Montroll 1953, ter Haar 1954 a, chap. 12). There are several models which are based on a more detailed picture of the ferromagnetic solids. First of all there is the original Heisenberg model (1928) or a generalization of it (Van Vleck 1953). In this model one uses the Heitler–London methods and the electrons (spins) responsible for the ferromagnetic behaviour are fixed in the lattice. Slater (1953; see also M.I.T. Solid State and Molecular Theory Group Technical Report, No. 6 (1954) which contains an extensive bibliography) prefers to link the theory of ferromagnetism to band theory (see also Zener and Heikes 1953 for a variant of the usual exchange interaction theories), while Stoner (see, for instance, Wohlfarth 1953) has developed the so-called collective electron theory of ferromagnetism. In those models the spins can wander through the lattice. At very low temperatures many people feel that Bloch's spin wave theory (1930) should give a fair picture of the situation. In order that this theory can be applied, the ground state of a ferromagnetic must be a singlet state of maximum multiplicity. Slater (see M.I.T. Report mentioned above) considers this unlikely to be realized and concludes "the spin-wave solution, elegant though it is, presumably has no bearing on the actual problem of ferromagnetism". Recently, however, data on neutron scattering in iron (Lowde 1954, 1956) have become available and these data cannot be explained on the basis of the collective electron model of ferromagnetism (Elliott 1956), but fit reasonably well with the behaviour predicted using spin-wave theory (Avakyants 1948, Moorhouse 1951, Elliott and Lowde 1955). Our discussion of ferromagnetism will be completely confined to the spin-wave theory which is valid at sufficiently low temperatures.

The basic idea of the spin-wave theory (Bloch 1930) is to describe the states of low excitation of a ferromagnetic in terms of ' wrong ' spins. The ground state is the one where all spins are parallel to all the others. The state function corresponding to this state is a linear combination of all state functions describing one ' wrong ' spin at different lattice sites, hence the name spin-waves. Bloch (1932) extended his model in a later paper in which he also introduced the so-called Bloch wall and Bethe (1931 ; see also Hulthén 1938) has shown that the exact low-lying energy levels of a linear chain of spins are the same as those obtained from the spin-wave theory. Kramers (1936 ; see also Opechowski 1937, Kramers and Heller 1934, Herring and Kittel 1951, and Keffer, Kaplan and Yafet 1953) has given a semi-classical discussion of spin-wave theory. In his first paper Bloch (1930) derived for the saturation magnetization $M(T)$ as a function of temperature the equation

$$M(T) = M(0)\,[1 - CT^{3/2}], \qquad \ldots\ldots(5.1)$$

where C is a constant. His result was derived for spin $S = \frac{1}{2}$, and Møller (1933) extended the discussion to arbitrary spin. Holstein and Primakoff (1940) derived the field dependence of the magnetization. Until recently there were no reliable estimates of the order of magnitude of the terms neglected in equation (5.1). Kramers (1936) and Opechowski (1937) found the next term to be proportional to T^2, while Schafroth (1954) and van Kranendonk (1955 a, b) found terms proportional to $T^{7/4}$. In a detailed and thorough investigation Dyson (1956 b) finds that the next term is proportional to $T^{5/2}$ and he finds, moreover, (Dyson 1956 a) that the interaction between spin-waves is much smaller than had been estimated earlier. This means that the spin-wave approximation is probably valid up to much higher temperatures than was previously thought. We must finally mention that the spin-wave theory can also be applied to antiferromagnetism (Anderson 1952, Ziman 1952 a, b, 1953 a, Kubo 1952, 1953). In this case it is more difficult to define mathematically the ground state, but if one assumes to know the ground state the rest of the analysis proceeds along lines similar to the ones employed in the case of ferromagnetism.

For the sake of simplicity we shall consider a linear chain of N spins, each pair being a distance a apart. Moreover, we shall only consider the case of total spin $S = \frac{1}{2}$ (compare the discussion in Keffer, Kaplan and Yafet 1953). The Hamiltonian will be of the form (compare ter Haar 1954 a, p. 255)

$$H = -2\mathscr{J} \sum_j (\mathbf{S}_j \, . \, \mathbf{S}_{j+1}), \qquad \ldots\ldots(5.2)$$

where $\mathscr{J}$ is the interaction energy, that is, the energy gained when two neighbours change from an antiparallel to a parallel alignment. In the case of ferromagnetism $\mathscr{J}$ is positive, and in the case of antiferromagnetism it is negative. The sum is over all spins in the chain, and $S_{N+1} = S_1$ is the periodicity condition. The energy $\mathscr{J}$ includes the factor $\hbar^2$ so that the components of the $\mathbf{S}_j$ are the Pauli matrices

$\sigma_x, \sigma_y, \sigma_z$ multiplied by $\frac{1}{2}$ (for the form of the Pauli matrices see, for instance, Kramers 1957, p. 272). In terms of the Pauli matrices, the Hamiltonian is

$$H = -\tfrac{1}{2}\mathcal{J}\sum_j [\sigma_{xj}\sigma_{xj+1}+\sigma_{yj}\sigma_{yj+1}+\sigma_{zj}\sigma_{zj+1}]. \qquad \ldots\ldots(5.3)$$

To find the eigenfunctions of the H we introduce the unit spinors ξ_i and η_i (see Kramers 1957, p. 238) and we have the following relations (e.g. Kramers 1957, p. 272)

$$\left.\begin{aligned} \sigma_{xj}\xi_j &= \eta_j, & \sigma_{yj}\xi_j &= i\eta_j, & \sigma_{zj}\xi_j &= \xi_j\,; \\ \sigma_{xj}\eta_j &= \xi_j, & \sigma_{yj}\eta_j &= -i\xi_j, & \sigma_{zj}\eta_j &= -\eta_j. \end{aligned}\right\} \qquad \ldots\ldots(5.4)$$

We see then easily that the following wave function will be an eigenfunction of H,

$$\psi_0 = \Pi_j\,\xi_j, \qquad \ldots\ldots(5.5)$$

corresponding to the eigenvalue

$$E_0 = -\tfrac{1}{4}N\mathcal{J}. \qquad \ldots\ldots(5.6)$$

In a ferromagnetic, the state described by ψ_0 is the ground state, where all spins are parallel. We can now consider excited states with one spin antiparallel; we shall call the wave function corresponding to the state where the l-th spin is reversed φ_l so that we have

$$\varphi_l = (\eta_l/\xi_l)\psi_0. \qquad \ldots\ldots(5.7)$$

This wave function is not an eigenfunction of H, but we can construct eigenfunctions by a linear superposition of the φ_l,

$$\psi = \Sigma c_l\varphi_l. \qquad \ldots\ldots(5.8)$$

The coefficients c_l can be obtained from the requirement that ψ is an eigenfunction of H. From equations (5.3), (5.4), (5.5), (5.7) and (5.8) we get

$$\begin{aligned} H\psi = (E\psi = E\Sigma c_l\varphi_l) &= -\mathcal{J}\sum_l c_l[\varphi_{l-1}+\varphi_{l+1}+\tfrac{1}{2}(N-4)\varphi_l] \\ &= -\mathcal{J}\sum_l [c_{l-1}+\tfrac{1}{2}(N-4)c_l+c_{l+1}]\varphi_l. \qquad \ldots\ldots(5.9) \end{aligned}$$

Equation (5.9) leads to a secular problem of a well-known form (for instance, Rutherford 1947, 1952) and is solved by the solution

$$c_l = A\exp(ikla). \qquad \ldots\ldots(5.10)$$

The parameter k is determined by the periodicity condition $c_l = c_{l+N}$, and we have (compare Slater 1930)

$$kNa/2\pi = 0, \quad \pm 1, \quad \pm 2, \ldots, \quad \pm\tfrac{1}{2}N. \qquad \ldots\ldots(5.11)$$

In equation (5.10) A is a normalization constant which is equal to $N^{-1/2}$.

We see that there are $2N+1$ different eigenfunctions corresponding to one wrong spin. The corresponding eigenvalues follow from equations (5.9) and (5.10),

$$E = -\mathcal{J}[\tfrac{1}{2}(N-4)+(c_{l-1}+c_{l+1})/c_l] = E_0+2\mathcal{J}(1-\cos ka). \ldots\ldots(5.12)$$

For small values of ka we have

$$E-E_0 \simeq \mathcal{J}a^2k^2. \qquad \ldots\ldots(5.13)$$

Let us consider for a moment the physical meaning of the wave functions (5.8) with c_l given by equation (5.10). We see that it corresponds to a wave where the amplitude of the ' wrongness ' of the spins at each point of the lattice varies sinusoidally with the position in the lattice. Hence the name ' spin-waves ' for these excitations. We note that the ψ are very closely related to the $\rho_{\mathbf{k}}$ of § 3.

If we had considered a three-dimensional lattice we would have found instead of equation (5.12) (for instance, Peierls 1955, p. 170)

$$E = E_0' + 2B \sum_l [1 - \cos(\mathbf{k}.\mathbf{a}_l)], \quad E_0' = -\tfrac{1}{4}Nz\mathcal{J}, \qquad \ldots\ldots(5.14)$$

where z is the number of nearest neighbours, where the summation is over the z lattice vectors $\mathbf{a}_l$ which join a given spin to its nearest neighbours, and where B is a constant proportional to $\mathcal{J}$. For small values of $|\mathbf{k}|$ we have

$$E - E_0' \simeq B \sum_l (\mathbf{k}.\mathbf{a}_l)^2 = Ba^2k^2, \qquad \ldots\ldots(5.15)$$

the last equality holding for cubic lattices of cube side a.

Up to now we have only considered states with one ' wrong ' spin. A basic assumption in the spin-wave theory—and one which seems well justified both by Dyson's analysis (1956 a, b) and by the experimental data both on neutron diffraction by iron mentioned above and on the intrinsic magnetization as a function of temperature (Fallot 1936)—is that states with two, three, ... ' wrong ' spins can be described as being a superposition of two, three, ... spin-waves. In that case the energy of the system will be given by the expression

$$E = \Sigma n_k \epsilon_k, \qquad \ldots\ldots(5.16)$$

where n_k is the number of spin-waves present corresponding to a wave number k, and where ϵ_k is given by equations (5.12) or (5.14). We see that the spin-waves are typical independent ' quasi-particles '. As n_k can be any integral number, the spin-waves are bosons, and the average value of n_k at a temperature T will be given by the equation

$$\overline{n_k} = [\exp(\beta\epsilon_k) - 1]^{-1}, \quad \beta = 1/kT. \qquad \ldots\ldots(5.17)$$

The simplest way to calculate the intrinsic magnetization is to calculate the total number of ' wrong ' spins at a given temperature, and we have thus

$$M(T) = M(0)[1 - 2\Sigma\overline{n_k}/N], \qquad \ldots\ldots(5.18)$$

where the summation is over all possible k-values, and where $\overline{n_k}$ is given by equation (5.17). The sum can be replaced by an integral and for sufficiently low temperatures we can use for ϵ_k equation (5.15), and we find

$$\Sigma\overline{n_k} \simeq (2\pi^2)^{-1} \int_0^\infty k^2\,dk\,[\exp(\beta Da^2k^2) - 1]^{-1} = cT^{3/2}, \qquad \ldots\ldots(5.19)$$

where the constant c depends on a, D, and on the type of lattice considered. Combining equations (5.18) and (5.19) we obtain equation (5.1).

The discussion in the antiferromagnetic case is more complicated and we refer to the papers mentioned at the beginning of this section.

§ 6. COLLECTIVE BEHAVIOUR OF NUCLEI

An extensive literature has been produced in the last few decades on the possible behaviour of nuclei. Possibly the first collective model of a nucleus was the liquid drop model introduced by Bohr (Bohr and Kalckar 1937) which was so very successful in the discussion of nuclear reactions and especially nuclear fission (Bohr and Wheeler 1939). On the other hand, the statistical theory of nuclear reactions (see, for instance, Weisskopf and Ewing 1940, Weisskopf 1947, ter Haar 1954 a, chap. 13) could be built up from an independent particle model to give the nuclear partition function. Even more disturbing was the success of the nuclear shell model (Mayer 1948) although Weisskopf (1951) has indicated how one can up to a point reconcile the liquid-drop–compound nucleus behaviour with the shell-independent particle behaviour (see § 1). In § 1 we mentioned briefly the success of the recent work by Brueckner, Eden, Bethe and others in explaining why a system of strongly interacting particles can also show shell structure. We do not want to pursue this topic here. There is, however, another aspect of the collective behaviour–individual particle dualism occurring in nuclei. In many respects one can analyse the collective behaviour of nuclei in terms of surface oscillations. In the next volume of the *Reports on Progress in Physics* Brink is going to discuss in detail the collective aspects of nuclei so that we can restrict ourselves to a discussion of an application of Tomonaga's method to the case of surface oscillations. We may refer here to papers by Bohr (1952, 1954), Bohr and Mottelson (1953), Coester (1955), Hill and Wheeler (1953), Süssmann (1954), and Tolhoek (1954, 1955) discussing in more detail collective nuclear models.

We use the formalism of § 2. Everything is simple, if we are considering only one mode of excitation. The generating function $\phi(\mathbf{x})$ may be given by the equation

$$\phi(\mathbf{x}) = r^l \mathrm{P}_l^m(\vartheta, \varphi), \qquad \ldots\ldots(6.1)$$

where P_l^m is a spherical harmonic. We use for the P_l^m the expressions given by Kramers (1957, p. 168). It is well known that the ξ derived from $\phi(\mathbf{x})$ by using equation (2.17) will be one of the velocity potentials occurring in the Bohr–Mottelson theory. We can illustrate the physical meaning of ξ by evaluating it in terms of the shape of the actual nuclear surface. We shall assume that we may replace the sum in equation (2.17) by an integral, so that we have

$$\xi = A^{-1}\int r^l \mathrm{P}_l^m(\vartheta, \varphi)\, d\mathbf{x}, \quad A = \int (\nabla\phi^* . \nabla\phi)\, d\mathbf{x}. \qquad \ldots\ldots(6.2)$$

In writing down equations (6.2) we have tacitly assumed that we are dealing with an incompressible homogeneous nucleus.

In evaluating ξ we introduce polar coordinates and we have

$$\xi = A^{-1}\int \mathrm{P}_l^m(\vartheta, \varphi)\, d\omega \int_0^R r^{l+2}\, dr, \qquad \ldots\ldots(6.3)$$

where $d\omega$ is an element of solid angle and where R is the distance of the actual nuclear surface from the centre. This means that R is a function of ϑ and φ,

$$R(\vartheta, \varphi) = R_0 + \Delta R_0(\vartheta, \varphi), \qquad \ldots\ldots(6.4)$$

where R_0 is the radius of the spherical, unexcited nucleus. Assuming ΔR_0 to be small compared with R_0 we have

$$\xi = (R_0^{l+2}/A)\int \Delta R_0(\vartheta,\varphi)\mathrm{P}_l^m(\vartheta,\varphi)\,d\boldsymbol{\omega}, \qquad \ldots\ldots(6.5)$$

which shows the relation between ξ and the collective parameters α_{lm} introduced by Bohr and Mottelson (1953, p. 10).

We shall now discuss the case where several modes are present at the same time. As in the recent paper by Miyazima and Tamura (1956) the discussion will be restricted to ξ's corresponding to $l=2$. The case $l=1$ leads to three ξ's which correspond to the centre-of-mass motion, while the case $l>2$ becomes rather too complicated for our present discussion. If $l=2$, there are five ξ's corresponding to $m=-2,-1,0,1,2$ which will be distinguished by an index m,

$$\xi_m = A_m^{-1}\int \phi_m\,d\mathbf{x}, \quad A_m = (-1)^m\int (\nabla\phi_{-m}\,.\,\nabla\phi_m)\,d\mathbf{x}, \quad \phi_m = r^2\mathrm{P}_2^{\,m}, \qquad \ldots\ldots(6.6)$$

where we have used the fact that $(-1)^m\phi_{-m}$ is the complex conjugate of ϕ_m.

Once again we are dealing with complex variables, and the conjugate momenta π_m are given by the equations

$$\pi_m = (-1)^m\int (\mathbf{p}\,.\,\nabla\phi_{-m})\,d\mathbf{x}. \qquad \ldots\ldots(6.7)$$

Let us now evaluate the various Poisson brackets. We are using for the P_l^m the normalization of Kramers (1957, p. 168) and we have then (Kramers 1957, p. 174)

$$\nabla_x\phi_m = r[\mathrm{P}_1^{\,m-1}-\mathrm{P}_1^{\,m+1}], \quad \nabla_y\phi_m = ir[\mathrm{P}_1^{\,m-1}+\mathrm{P}_1^{\,m+1}], \quad \nabla_z\phi_m = 2r\mathrm{P}_1^{\,m}. \qquad \ldots\ldots(6.8)$$

The Poisson brackets $\{\xi_m,\xi_{m'}\}$ are equal to zero, while we have for the other brackets

$$\{\pi_m,\xi_{m'}\} = (-1)^{m+1}A_{m'}^{-1}\int (\nabla\phi_{-m}\,.\,\nabla\phi_{m'})\,d\mathbf{x}, \qquad \ldots\ldots(6.9)$$

$$\{\pi_m,\pi_{m'}\} = (-1)^{m+m'}\sum_{i,j}\int p_i[\nabla_i\nabla_j\phi_{-m'}\nabla_j\phi_{-m} - \nabla_i\nabla_j\phi_{-m}\nabla_j\phi_{-m'}]\,d\mathbf{x}, \qquad \ldots\ldots(6.10)$$

where i,j run through x,y,z. We see that $\{\pi_m,\xi_m\} = -1$ as would be expected (see § 2), while for $m \neq m'$ we get using equations (6.8)

$$\begin{aligned}\{\pi_m,\xi_{m'}\} &= 2(-1)^m A_{m'}^{-1}\int r^2[\mathrm{P}_1^{\,-m'+1}\mathrm{P}_1^{\,m-1}+\mathrm{P}_1^{\,-m'-1}\mathrm{P}_1^{\,m+1}-2\mathrm{P}_1^{\,-m'}\mathrm{P}_1^{\,m}]\,d\mathbf{x}\\ &= -\delta_{mm'}, \qquad \ldots\ldots(6.11)\end{aligned}$$

if we neglect terms of the order $\Delta R_0/R_0$ as we shall do.

Similarly we find for the $\{\pi_m,\pi_{m'}\}$

$$\left.\begin{aligned}\{\pi_m,\pi_m\} = 0, \quad \{\pi_m,\pi_{m'}\} &= (-1)^{m+m'}\{\pi_{-m},\pi_{-m'}\}^*,\\ \{\pi_2,\pi_1\} = \{\pi_2,\pi_0\} &= 0,\\ \{\pi_2,\pi_{-1}\} = iM_x - M_y, \quad \{\pi_2,\pi_{-2}\} &= -iM_z,\\ \{\pi_1,\pi_0\} = -3iM_x+3M_y, \quad \{\pi_1,\pi_{-1}\} &= -2iM_z,\end{aligned}\right\} \qquad \ldots\ldots(6.12)$$

where we have used (i) equations (6.8), (ii) the equations

$$\nabla_x r\mathrm{P}_1{}^m = \tfrac{1}{2}(\mathrm{P}_0{}^{m-1} - \mathrm{P}_0{}^{m+1}), \quad \nabla_y r\mathrm{P}_1{}^m = \tfrac{1}{2}i(\mathrm{P}_0{}^{m-1} + \mathrm{P}_0{}^{m+1}), \quad \nabla_z r\mathrm{P}_1{}^m = \mathrm{P}_0{}^m, \qquad \ldots\ldots(6.13)$$

and (iii) the following expressions for the $\mathrm{P}_l{}^m$ (see, for instance, Kramers 1957, § 45)

$$\mathrm{P}_l{}^m = 0, \quad |m| > l; \qquad \mathrm{P}_0{}^0 = 1; \qquad r\mathrm{P}_1{}^{\pm 1} = \tfrac{1}{2}iy \pm \tfrac{1}{2}x, \quad r\mathrm{P}_1{}^0 = z. \qquad \ldots\ldots(6.14)$$

Finally, $\mathbf{M}$ is the total angular momentum vector,

$$\mathbf{M} = \int [\mathbf{x} \times \mathbf{p}]\, d\mathbf{x}. \qquad \ldots\ldots(6.15)$$

We note that, if the total angular momentum of the nucleus vanishes, the π_m, ξ_m constitute a satisfactory set of canonical coordinates. This requirement of vanishing total angular momentum can be compared to the requirement of vanishing total linear momentum in § 3.

If $l > 2$ the situation becomes more complicated, and in the expressions for the $\{\pi_m, \pi_{m'}\}$ we get integrals of the kind $\int x^a y^b z^c p_i \, d\mathbf{x}$, where $a+b+c = l-1$, and where p_i is one of the components of $\mathbf{p}$.

We may remark here that Miyazima and Tamura (1956) have also discussed the $l = 2$ case. They use quantum mechanical formulae and they, moreover, use subsidiary conditions as was done by Bohm and Pines in the case of plasma oscillations. This was necessary as they did not make the assumption of zero angular momentum.

We could now use the ξ_m and π_m to derive by the methods discussed in § 2 an expression for the Hamiltonian and we could then compare the resultant Hamiltonian with Bohr and Mottelson's Hamiltonian. The calculations are, however, very complicated and would lead us too far afield.

§ 7. LIQUID HELIUM

The behaviour of liquid helium has presented a strong challenge to both theorists and experimentalists. The only substance to remain liquid down to the absolute zero because of its large zero point motion, it behaves in a most unexpected manner by showing below its λ-point at 2·2°K curious superfluid phenomena. As the forces between helium atoms can be derived more or less satisfactorily from first principles and since they are very weak, it was hoped to produce a theory which could explain the peculiar behaviour of He II from first principles. As the forces are so very weak, many authors believed that the phenomenon of the condensation of a perfect Bose–Einstein gas (Einstein 1925, London 1938, Fowler and Jones 1938; for a more extensive bibliography see, for instance, ter Haar 1954 a, p. 224) would be essentially the explanation of the λ-transition. Feynman (1953 a; see also Chester 1954 a, b) made it plausible that the weak interatomic forces would not alter this conclusion, but the present author (ter Haar 1954 b; see also 1954 a, p. 215) has expressed doubts as to the validity of Feynman's argument. Be that as it may, there is very little doubt that the behaviour of liquid helium near the absolute zero is mainly determined by its statistics, and this will be the subject of our discussion in this section.

One can express many properties of liquid helium in terms of its energy spectrum, that is, in terms of the energy, $\epsilon(k)$, as a function of the wave number k. From a quantization of the hydrodynamic equations Landau (1941, 1947) deduced for $\epsilon(k)$ the following form

$$\epsilon(k) = \hbar ck, \quad k \ll k_0\,; \qquad \epsilon(k) = \Delta + \hbar^2(k-k_0)^2/2\mu, \quad k \sim k_0, \qquad \ldots\ldots(7.1)$$

where c, Δ, k_0 and μ are parameters. This energy spectrum shape is sufficient to explain the observed second sound velocity (Osborne 1956) and the observed specific heat data (Kramers, Wasscher and Gorter 1952). Feynman (1953 b, 1954, 1955, Feynman and Cohen 1956 ; see also Bogoliubov and Zubarev 1955) has been able to derive the spectrum (7.1) from first principles, while Brenig (1956) has shown how Tomonaga's method can be adapted to obtain the same result. We shall discuss Brenig's method and we refer to a survey by Wilks (1957) in this same volume for a general discussion of Feynman's work. We may mention here that Kronig and Thellung (1952), Thellung (1953, 1956), Ziman (1953 b), and Allcock and Kuper (1955) have tried to obtain the spectrum (7.1) by quantizing the hydrodynamical equations of motion of a fluid.

For our discussion it is essential (i) that here the *low*-lying excitations are phonons, that is, can be described by using the $\rho_{\mathbf{k}}$ of § 3, and that it thus takes a finite amount of energy to excite the rotons or particle modes, and (ii) that the interaction between helium atoms is so small that it can be neglected. We shall start our discussion using classical formulae in order that we may see clearly what would be suitable coordinates to describe the system, but in order to find the energy spectrum we shall have to go over to quantum mechanics.

We shall normalize the Fourier components of the density slightly differently from how we did it in § 3. We shall use the number density rather than the mass density so that we have

$$n(\mathbf{x}) = \sum_j \delta(\mathbf{x}-\mathbf{x}_j), \qquad \ldots\ldots(7.2)$$

and we introduce instead of the $\rho_{\mathbf{k}}$ the $n_{\mathbf{k}}$ given by the equations

$$n_{\mathbf{k}} = \int n(\mathbf{x}) \exp[i(\mathbf{k}.\mathbf{x})]\,d\mathbf{x}, \quad \text{or,} \qquad \ldots\ldots(7.3)$$

$$n_{\mathbf{k}} = \sum_j \exp[i(\mathbf{k}.\mathbf{x}_j)], \quad n_{-\mathbf{k}} = \sum_j \exp[-i(\mathbf{k}.\mathbf{x}_j)]. \qquad \ldots\ldots(7.4)$$

The collective momenta $\pi_{\pm\mathbf{k}}$ are given by the equations [compare eqn (3.10)]

$$\pi_{\pm\mathbf{k}} = \pm(iNk^2)^{-1}\sum_j (\mathbf{p}_j.\mathbf{k})\exp[\mp i(\mathbf{k}.\mathbf{x}_j)], \qquad \ldots\ldots(7.5)$$

and once again we have

$$\{n_{\pm\mathbf{k}}, n_{\pm\mathbf{k}'}\} = \{\pi_{\pm\mathbf{k}}, \pi_{\pm\mathbf{k}'}\} = \{n_{+\mathbf{k}}, n_{-\mathbf{k}'}\} = \{\pi_{+\mathbf{k}}, \pi_{-\mathbf{k}'}\} = 0, \qquad \ldots\ldots(7.6)$$

$$\{\pi_{+\mathbf{k}}, n_{+\mathbf{k}'}\} = \{\pi_{-\mathbf{k}}, n_{-\mathbf{k}'}\} = -\delta_{\mathbf{k}\mathbf{k}'}, \quad \{\pi_{+\mathbf{k}}, n_{-\mathbf{k}'}\} = \{\pi_{-\mathbf{k}}, n_{+\mathbf{k}}\} = 0, \qquad \ldots\ldots(7.7)$$

where we have assumed $\Sigma\mathbf{p}_j$ to be zero, and where we have neglected terms of the order $N^{-1/2}$. The collective kinetic energy will be of the form

$$T_c = \sum_{\mathbf{k}} (\pi_{+\mathbf{k}}\pi_{-\mathbf{k}}/2mNk^2). \qquad \ldots\ldots(7.8)$$

The collective potential energy can most easily be obtained from the phenomenological expressions for the potential energy in a sound wave,

$$U = \tfrac{1}{2}mc^2 \int [n(\mathbf{x}) - n_0]^2 \, d\mathbf{x} = \tfrac{1}{2}\Sigma \mu_{\mathbf{k}} \omega_{\mathbf{k}}^2 n_{\mathbf{k}} n_{-\mathbf{k}}, \qquad \ldots\ldots(7.9)$$

where c is the sound velocity, n_0 is the equilibrium number density, $n_0 = N/V$, with N the total number of particles and V the volume available, and $\mu_{\mathbf{k}}$ and $\omega_{\mathbf{k}}$ are given by the equations

$$\mu_{\mathbf{k}} = m/Nk^2, \quad \omega_{\mathbf{k}} = ck. \qquad \ldots\ldots(7.10)$$

We get the following collective Hamiltonian

$$H_c = \sum_{\mathbf{k}} \{(\pi_{+\mathbf{k}} \pi_{-\mathbf{k}}/2mNk^2) + (mc^2 n_{+\mathbf{k}} n_{-\mathbf{k}}/N)\}, \qquad \ldots\ldots(7.11)$$

which in the well-known way (see, for instance, Penrose 1954) leads to the following collective energy levels

$$E_{c\mathbf{k}} = \hbar ck. \qquad \ldots\ldots(7.12)$$

In order to introduce particle coordinates we first of all note that we can only use the $n_{\mathbf{k}}$ for small values of $|\mathbf{k}|$, say $|\mathbf{k}| < k_D$ corresponding to the Debye cut-off in the theory of lattice vibrations (Debye 1912). Secondly, as the $n_{\mathbf{k}}$ correspond only to longitudinal waves, we should expect the first N excitations to correspond to the $n_{\mathbf{k}}$, which means that k_D should be given by the equation

$$k_D^3 = 6\pi^2 n_0. \qquad \ldots\ldots(7.13)$$

This assumption is probably not too bad a one, as the specific heat for sufficiently low T follows, indeed, a T^3 law; it means essentially the neglect of dispersion.

Before introducing the $2N$ particle coordinates which together with the N $n_{\mathbf{k}}$ will take care of the $3N$ degrees of freedom of our system, we note that the displacements $\mathbf{s}_j$ of the particles from their equilibrium positions $\mathbf{X}_j$,

$$\mathbf{x}_j = \mathbf{X}_j + \mathbf{s}_j(\mathbf{X}) \simeq \mathbf{X}_j + \mathbf{s}_j(\mathbf{x}), \qquad \ldots\ldots(7.14)$$

can be expressed in terms of the $n_{\mathbf{k}}$ by using the relation

$$\sum_{\mathbf{k}} n_{\mathbf{k}} \exp[-i(\mathbf{k} . \mathbf{X})] = n_0, \qquad \ldots\ldots(7.15)$$

which is satisfied by the following expression for the $\mathbf{s}_j$ [provided they are so small that we may write $\exp\{-i(\mathbf{k} . \mathbf{s})\} \simeq 1 - i(\mathbf{k} . \mathbf{s})$]

$$\mathbf{s}_j = \sum_{\mathbf{k}} (i\mathbf{k}/Nk^2) n_{\mathbf{k}} \exp[-i(\mathbf{k} . \mathbf{x}_j)]. \qquad \ldots\ldots(7.16)$$

The $\mathbf{X}_j$ are not independent, as they satisfy the equations

$$\begin{aligned} N\sum_j \exp[i(\mathbf{k} . \mathbf{X}_j)] &\simeq N\sum_j \exp[i(\mathbf{k} . \mathbf{x}_j)]\{1 - i(\mathbf{k} . \mathbf{s}_j)\} \\ &= N\sum_j \exp[i(\mathbf{k} . \mathbf{x}_j)] - \sum_{\mathbf{k}'} [(\mathbf{k} . \mathbf{k}')/k'^2] n_{\mathbf{k}'} n_{\mathbf{k}-\mathbf{k}'} = 0, \qquad \ldots\ldots(7.17) \end{aligned}$$

if we use once again the random phase approximation, which means in this case retaining only the term with $\mathbf{k}' = \mathbf{k}$ in the sum over $\mathbf{k}'$. The N relations (7.17) for the N possible values of $\mathbf{k}$ reduce the number of independent coordinates $\mathbf{X}_j$ to $2N$ which together with the N $n_{\mathbf{k}}$ form the complete set of coordinates.

The $\mathbf{X}_j$ are orthogonal to the $n_{\mathbf{k}}$ to the same approximation as equations (7.17) are satisfied, since ($\mathbf{X}_{j\kappa}$ is the κth component of $\mathbf{X}_j$)

$$\sum_m (\nabla_m \mathbf{X}_{j\kappa} . \nabla_m n_{\mathbf{k}}) = i\Big\{k_\kappa + \sum_{\mathbf{k}'} [k_\kappa'(\mathbf{k}.\mathbf{k}')/k'^2]$$
$$\times (n_{\mathbf{k}'} \exp[-i(\mathbf{k}'.\mathbf{x}_j)] - n_{\mathbf{k}-\mathbf{k}'} \exp[-i(\mathbf{k}-\mathbf{k}'.\mathbf{x}_j)])\Big\} \exp[i(\mathbf{k}.\mathbf{x}_j)] \simeq 0. \qquad \ldots\ldots(7.18)$$

From Feynman's considerations we know that it is likely that it will take a finite amount of energy to excite the particle excitations. This means that we may solve the Schrödinger equation using the Born–Oppenheimer adiabatic approximation, or we may, in determining the $\mathbf{X}_j$ dependence of the total wave function, to a first approximation neglect T_c and consider the $n_{\mathbf{k}}$ to have fixed values $\overline{n_{\mathbf{k}}}$. The differential equation for the $\mathbf{X}_j$ dependence of the wave function reduces in this approximation to the form

$$(H - T_c)\varphi(\mathbf{X}_j, \overline{n_{\mathbf{k}}}) = E(\overline{n_{\mathbf{k}}})\varphi(\mathbf{X}_j, \overline{n_{\mathbf{k}}}) \qquad \ldots\ldots(7.19)$$

and the total wave function Φ including $n_{\mathbf{k}}$ and $\mathbf{X}_j$ will be of the form

$$\Phi = \psi(n_{\mathbf{k}})\varphi(\mathbf{X}_j, \overline{n_{\mathbf{k}}}), \qquad \ldots\ldots(7.20)$$

and will satisfy the Schrödinger equation

$$H\Phi = E\Phi, \qquad \ldots\ldots(7.21)$$

where H in equations (7.19) and (7.21) is the total Hamiltonian of the system. We assume the $\overline{n_{\mathbf{k}}}$ to be small so that we have

$$\varphi(\mathbf{X}_j, \overline{n_{\mathbf{k}}}) \simeq \varphi(\mathbf{X}_j, 0) + \varphi'(\mathbf{X}_j, 0)\overline{n_{\mathbf{k}}}, \qquad \ldots\ldots(7.22)$$

$$E(\overline{n_{\mathbf{k}}}) \simeq E(0) + \tfrac{1}{2}\Sigma E_{\mathbf{k}}''(0)\overline{n_{\mathbf{k}}}\,\overline{n_{-\mathbf{k}}}. \qquad \ldots\ldots(7.23)$$

(We are using here the well-known result from perturbation theory that the second-order terms in $E(\overline{n_{\mathbf{k}}})$ correspond to first-order terms in φ; there is no first-order term in E, because $\overline{n_{\mathbf{k}}} = 0$ corresponds to the equilibrium configuration.) One can show (see Brenig 1956) that equations (7.19) to (7.23) lead to the following equation for $\psi(n_{\mathbf{k}})$,

$$H_c\psi = [E - E(0)]\psi. \qquad \ldots\ldots(7.24)$$

Equation (7.19) for φ can be written in the form

$$(H - T_c)\chi = (H + U_{\text{eff}})\chi = E(0)\chi, \qquad \ldots\ldots(7.25)$$

where we have neglected the second-order terms in $E(\overline{n_{\mathbf{k}}})$, where U_{eff} is an effective potential energy,

$$U_{\text{eff}} = -[T_c\Delta(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}})]/\Delta(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}}), \qquad \ldots\ldots(7.26)$$

and χ is an effective wave function

$$\chi = \varphi\Delta(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}}), \qquad \ldots\ldots(7.27)$$

while $\Delta(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}})$ is given by the equation

$$\Delta(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}}) = \pi^{-1/2}\Pi_{\mathbf{k}}\alpha_{\mathbf{k}}^{-1}\exp[-\alpha_{\mathbf{k}}^{-2}(n_{\mathbf{k}} - \overline{n_{\mathbf{k}}})^2]. \qquad \ldots\ldots(7.28)$$

Brenig calls U_{eff} a ' Zwangspotential '; it sees to it that the individual motions of the particles are such that $n_{\mathbf{k}}$ is equal to $\overline{n_{\mathbf{k}}}$. However, as the separation between particle modes and collective modes is not rigorous, $n_{\mathbf{k}}$ can only approximately be equal to $\overline{n_{\mathbf{k}}}$, hence the Gaussian form of Δ instead of a proper delta-function. Of course, the final results one obtains by using U_{eff} should be independent of the width of the Gaussian, that is, of the $\alpha_{\mathbf{k}}$. Use of χ instead of φ means that we do not have to worry about selecting carefully only $2N$ from among the $3N\mathbf{X}_j$. This means that the introduction of the Δ-functions takes the place of the subsidiary conditions of Bohm and Pines (1953).

Equation (7.25) can now be used to evaluate $E(0)$, that is, the particle part of the total energy. The calculations are rather complicated and we refer to Brenig (1956) for a detailed discussion. The final result is that $E(0)$ is given by the equation

$$\left.\begin{array}{l} E(0) = \sum\limits_{\mathbf{k}} \epsilon(\mathbf{k}) N_{\mathbf{k}}, \quad N_0 + \sum\limits_{|\mathbf{k}| > k_{\mathrm{D}}} N_{\mathbf{k}} = \frac{2}{3}N\,; \\ \epsilon(\mathbf{k}) = 0, \quad k < k_{\mathrm{D}}, \quad \epsilon(\mathbf{k}) = \hbar^2 k^2/2m^*, \quad k > k_{\mathrm{D}}, \quad m^* = \frac{3}{2}m\,; \end{array}\right\} \quad \ldots\ldots(7.29)$$

that is, the spectrum of a perfect Bose–Einstein gas of $\frac{2}{3}N$ particles and an effective mass $\frac{3}{2}m$.

ACKNOWLEDGMENTS

The author would like to express his thanks to all who took part in a symposium on collective behaviour held at Harwell during the summer of 1956. The discussions at that symposium greatly helped him to clarify his ideas on several points of the preceding text.

REFERENCES

Allcock, G. R., and Kuper, C. G., 1955, *Proc. Roy. Soc.* A, **231**, 226.

Anderson, P. W., 1952, *Phys. Rev.*, **86**, 694.

Anselm, A. I., 1951, *J. Tech. Phys., USSR*, **21**, 489.

Avakyants, G., 1948, *J. Exp. Theor. Phys., USSR*, **18**, 444.

Bardeen, J., 1950 a, *Phys. Rev.*, **79**, 167 ; 1950 b, *Ibid.*, **80**, 567 ; 1951 a, *Ibid.*, **81**, 829 ; 1951 b, *Rev. Mod. Phys.*, **23**, 261.

Bethe, H. A., 1931, *Z. Phys.*, **71**, 205 ; 1956, *Phys. Rev.*, **103**, 1353.

Bloch, F., 1929, *Z. Phys.*, **57**, 545 ; 1930, *Ibid.*, **61**, 206 ; 1932, *Ibid.*, **74**, 295 ; 1934, *Helv. Phys. Acta*, **7**, 385.

de Boer, J., 1949, *Rep. Progr. Phys.*, **12**, 305 (London : Physical Society).

de Boer, J., and Michels, A., 1939, *Physica*, **6**, 97.

Bogoliubov, N. N., and Zubarev, D. N., 1955, *J. Exp. Theor. Phys., USSR*, **28**, 129 (=*Soviet Phys., JETP*, **1**, 83).

Bohm, D., and Gross, E. P., 1949 a, b, *Phys. Rev.*, **75**, 1851, 1865.

Bohm, D., and Pines, D., 1951, *Phys. Rev.*, **82**, 625 ; 1953, *Ibid.*, **92**, 609.

Bohr, A., 1952, *K. Danske Vidensk. Selsk., Nat.-Fys. Medd.* **26**, No. 14 ; 1954, *Rotational States of Atomic Nuclei* (Copenhagen : Munksgaard).

Bohr, A., and Mottelson, B. R., 1953, *K. Danske Vidensk. Selsk., Nat.-Fys. Medd.*, **27**, No. 16.

Bohr, N., and Kalckar, F., 1937, *K. Danske Vidensk. Selsk., Nat.-Fys. Medd.* **14**, No. 10.

Bohr, N., and Wheeler, J. A., 1939, *Phys. Rev.*, **56**, 426.

Bontsch-Brujewitsch, W. L. (=Bonch-Bruevich, V. L.), 1955, *Fortschr. Phys.*, **3**, 371 ; 1956, *J. Exp. Theor. Phys., USSR*, **30**, 342 (=*Soviet Phys., JETP*, **3**, 278).

Born, M., and Green, H. S., 1949, *A General Kinetic Theory of Liquids* (Cambridge : University Press).

Born, M., and von Kármán, T., 1912, *Phys. Z.*, **13**, 297 ; 1913, *Ibid.*, **14**, 15.

Brenig, W., 1956, *Z. Phys.*, **144**, 488.

Brussels, 1957, *Proceedings of the* 1956 *Statistical Mechanics Conference* (New York : Interscience).

Chester, G. V., 1954 a, *Phys. Rev.*, **93**, 1412 ; 1954 b, *Ibid.*, **94**, 246.

Clemmow, P. C., and Willson, A. J., 1956, *Proc. Roy. Soc.* A, **237**, 117.

Coester, F., 1955, *Phys. Rev.*, **99**, 170.

Debye, P., 1912, *Ann. Phys., Lpz.*, **39**, 789.

Debye, P., and Hückel, E., 1923, *Phys. Z.*, **24**, 185.

Dyson, F. J., 1956 a, b, *Phys. Rev.*, **102**, 1217, 1230.

Ecker, G., 1955 a, b, *Z. Phys.*, **140**, 274, 293 ; 1955 c, *Ibid.*, **141**, 294 ; 1955 d, *Appl. Sci. Res.* B, **5**, 321.

Eden, R. J., 1956, *Proc. Roy. Soc.* A, **235**, 408.

Einstein, A., 1925, *Berliner Ber.*, 3.

Elliott, R. J., 1956, *Proc. Roy. Soc.* A, **235**, 289.

Elliott, R. J., and Lowde, R. D., 1955, *Proc. Roy. Soc.* A, **230**, 46.

Elsasser, W., 1933, *J. Phys. Radium*, **4**, 549 ; 1934 a, b, *Ibid.*, **5**, 389, 635.

Fallot, M., 1936, *Ann. Phys., Paris*, **6**, 305.

Feynman, R. P., 1953 a, b, *Phys. Rev.*, **91**, 1291, 1301 ; 1954, *Ibid.*, **94**, 262 ; 1955, *Progr. Low Temp. Phys.*, **1**, 17.

Feynman, R. P., and Cohen, M., 1956, *Phys. Rev.*, **102**, 1189.

Fock, V., 1930, *Z. Phys.*, **61**, 126.

Fowler, R. H., and Jones, H., 1938, *Proc. Camb. Phil. Soc.*, **34**, 573.

Fröhlich, H., 1950, *Phys. Rev.*, **79**, 845 ; 1954, *Physica*, **19**, 755.

Galasiewicz, Z., 1955, *Acta Phys. Polon.*, **14**, 373 ; 1956 a, b, *Ibid.*, **15**, 49, 79.

Gibbs, J. W., 1902, *Elementary Principles in Statistical Mechanics* (New York : Scribner).

Gombas, P., 1949, *Die statistische Theorie des Atoms und ihre Anwendungen* (Vienna : Springer) ; 1954, *Acta Phys. Hung.*, **4**, 187 ; 1956 a, b, *Ibid.*, **5**, 483, 503.

Gombas, P., and Kunvári, O., 1955, *Acta Phys. Hung.*, **5**, 339.

Gombas, P., and Ladányi, K., 1955, *Acta Phys. Hung.*, **5**, 313.

Gombas, P., and Szépfalusy, P., 1955, *Acta Phys. Hung.*, **5**, 259.

Green, H. S., 1952, *A Molecular Theory of Fluids* (Amsterdam : North-Holland).

Groschwitz, E., and Siebertz, K., 1956, *Z. Naturf.*, **11a**, 482.

Guggenheimer, K. M., 1934 a, b, *J. Phys. Radium*, **5**, 253, 475.

ter Haar, D., 1954 a, *Elements of Statistical Mechanics* (New York : Rinehart) ; 1954 b, *Phys. Rev.*, **95**, 895.

Hartree, D. R., 1928, *Proc. Camb. Phil. Soc.*, **24**, 89.

Heisenberg, W., 1928, *Z. Phys.*, **49**, 619.

Herring, C., and Kittel, C., 1951, *Phys. Rev.*, **81**, 869.

Hill, D. L., and Wheeler, J. A., 1953, *Phys. Rev.*, **89**, 1102.

Hirschfelder, J. O., Curtiss, C. F., and Bird, R. B., 1954, *Molecular Theory of Gases and Liquids* (New York : Wiley).

Holstein, T., and Primakoff, H., 1940, *Phys. Rev.*, **58**, 1098.

Hubbard, J., 1954, *Proc. Phys. Soc.* A, **67**, 1058.

Hulthén, L., 1938, *Ark. Mat. Astron. Fys.*, **26A**, No. 11.

Ising, E., 1925, *Z. Phys.*, **31**, 253.

Jastrow, R., 1955, *Phys. Rev.*, **98**, 1479.

Kahn, B., 1938, *Thesis* (University of Utrecht).

Kahn, B., and Uhlenbeck, G. E., 1938, *Physica*, **5**, 399.

van Kampen, N. G., 1955, *Physica*, **21**, 949.

Kanazawa, H., 1955, *Progr. Theor. Phys., Japan*, **13**, 227.

Keffer, F., Kaplan, H., and Yafet, Y., 1953, *Amer. J. Phys.*, **21**, 250.

Kinoshita, T., and Nambu, Y., 1954, *Phys. Rev.*, **94**, 598.

Kirkwood, J. G., 1946, *J. Chem. Phys.*, **14**, 180.

Kittel, C., 1953, *Introduction to Solid State Physics* (New York : Wiley).

Kramers, H. A., 1936, *Leiden Commun.*, **22**, Suppl. No. 83 ; 1938, *Nuovo Cim.*, **15**, 108 ; 1957, *Quantum Mechanics* (Amsterdam : North-Holland).

Kramers, H. A., and Heller, G., 1934, *Proc. Acad. Sci. Amst.*, **37**, 378.

Kramers, H. C., Wasscher, J. D., and Gorter, C. J., 1952, *Physica*, **18**, 329.

van Kranendonk, J., 1955 a, b, *Physica*, **21**, 749, 925.

Kronig, R. L. de L., and Thellung, A., 1952, *Physica*, **18**, 749.

Kubo, R., 1952, *Phys. Rev.*, **86**, 929 ; 1953, *Rev. Mod. Phys.*, **25**, 344.

Lamb, W. E., Jr., and Retherford, R. C., 1947, *Phys. Rev.*, **72**, 241.

Landau, L., 1941, *J. Phys. USSR*, **5**, 71 ; 1946, *Ibid.*, **10**, 25 ; 1947, *Ibid.*, **11**, 91.

London, F., 1938, *Phys. Rev.*, **54**, 947.

Lorentz, H. A., 1909, *The Theory of Electrons* (Leipzig : Teubner).

Lowde, R. D., 1954, *Proc. Roy. Soc.* A, **221**, 206 ; 1956, *Ibid.*, **235**, 305.

McGinnies, R. T., and Jansen, L., 1956, *Phys. Rev.*, **101**, 1301.

Mayer, J. E., 1937, *J. Chem. Phys.*, **5**, 67.

Mayer, J. E., and Mayer, M. G., 1940, *Statistical Mechanics* (New York : Wiley).

Mayer, M. G., 1948, *Phys. Rev.*, **74**, 235.

Meyer, H. J. G., 1956 a, *Physica*, **22**, 109 ; 1956 b, *Thesis* (University of Amsterdam).

Miyazima, T., and Tamura, T., 1956, *Progr. Theor. Phys., Japan*, **15**, 255.

Møller, C., 1933, *Z. Phys.*, **82**, 559.

Montroll, E. W., and Mayer, J. E., 1941, *J. Chem. Phys.*, **9**, 626.

Moorhouse, R. G., 1951, *Proc. Phys. Soc.* A, **64**, 1097.

Morse, P. M., 1936, *Vibration and Sound* (New York : McGraw-Hill).

Mott, N. F., 1954, *Proc. 10th Solvay Congress.*

Newell, G. F., and Montroll, E. W., 1953, *Rev. Mod. Phys.*, **25**, 353.

Nix, F. C., and Shockley, W., 1938, *Rev. Mod. Phys.*, **10**, 1.

Onsager, L., 1944, *Phys. Rev.*, **65**, 117.

Opechowski, W., 1937, *Physica*, **4**, 715.

Osborne, D. V., 1956, *Phil. Mag.*, **1**, 301.

Peierls, R. E., 1955, *The Quantum Theory of Solids* (Oxford : University Press).

Pekar, S. I., 1946, *J. Exp. Theor. Phys. USSR*, **16**, 341 ; 1954 a, *Untersuchungen über die Elektronentheorie der Kristalle* (Berlin : Akademie Verlag) ; 1954 b, *Fortschr. Phys.*, **1**, 367.

Penrose, O., 1954, *Phil. Mag.*, **45**, 80.

Percus, J. K., and Yevick, G. J., 1956, *Phys. Rev.*, **101**, 1192.

Pines, D., 1953, *Phys. Rev.*, **92**, 626 ; 1954, *Proc. 10th Solvay Congress* ; 1955, *Solid State Phys.*, **1**, 367 ; 1956, *Rev. Mod. Phys.*, **28**, 184.

Pines, D., and Bohm, D., 1952, *Phys. Rev.*, **85**, 338.

Poincaré, H., 1892, *Les méthodes nouvelles de la mécanique céleste* (Paris : Gauthiers-Villars).

Pomerantschuk, I. J., 1941 a, b, *J. Exp. Theor. Phys. USSR*, **11**, 226, 246.

Pomerantschuk, I. J., and Achieser, A. I., 1944, *J. Exp. Theor. Phys. USSR*, **14**, 342.

Raimes, S., 1957, *Rep. Progr. Phys.*, **20**, 1 (London : Physical Society).

Riesenfeld, W. B., and Watson, K. M., 1956, *Phys. Rev.*, **104**, 492.

Rutherford, D. E., 1947, *Proc. Roy. Soc. Edinb.* A, **62**, 229 ; 1952, *Ibid.*, **63**, 232.

Salam, A., 1954, *Progr. Theor. Phys., Japan*, **9**, 550.

Schafroth, M. R., 1951, *Helv. Phys. Acta*, **24**, 645 ; 1952, *Nuovo Cim.*, **9**, 291 ; 1954, *Proc. Phys. Soc.* A, **67**, 33.

Schiff, L. I., 1949, *Quantum Mechanics* (New York : McGraw-Hill).

Seitz, F., 1940, *The Modern Theory of Solids* (New York : McGraw-Hill).

Skinner, R., 1956, *Canad. J. Phys.*, **34**, 901.

Skyrme, T. H. R., 1957, *Proc. Roy. Soc.* A, **239**, 399.

Slater, J. C., 1929, *Phys. Rev.*, **34**, 1293 ; 1930, *Ibid.*, **35**, 509 ; 1951 a, *J. Chem. Phys.*, **19**, 220 ; 1951 b, *Phys. Rev.*, **82**, 538 ; 1953, *Rev. Mod. Phys.*, **25**, 199.

Sommerfeld, A., 1928, *Z. Phys.*, **47**, 1.

Sommerfeld, A., and Bethe, H., 1933, *Handb. Phys.*, $\mathbf{24_2}$, 333.

Steenbeck, M., 1932, *Z. Phys.*, **76**, 260.

Stoner, E. C., 1948, *Rep. Progr. Phys.*, **11**, 43 (London : Physical Society).

Süssmann, G., 1954, *Z. Phys.*, **139**, 543.

Thellung, A., 1953, *Physica*, **19**, 217 ; 1956, *Helv. Phys. Acta*, **29**, 103.

Thomson, J. J., and Thomson, G. P., 1933, *Conduction of Electricity through Gases*, Vol. **2** (Cambridge: University Press), p. 356.

Tisza, L., 1950, *Phys. Rev.*, **80**, 717.

Tolhoek, H. A., 1954, *Ned. Tijdschr. Natuurk.*, **20**, 259; 1955, *Physica*, **21**, 1.

Tomonaga, 1950, *Progr. Theor. Phys., Japan*, **5**, 544; 1955 a, b, *Ibid.*, **13**, 467, 482.

Tonks, L., and Langmuir, I., 1929, *Phys. Rev.*, **33**, 195.

Van Vleck, J. H., 1953, *Rev. Mod. Phys.*, **25**, 220.

Vlasov, A., 1945, *J. Phys. USSR*, **9**, 25.

Weisskopf, V. F., 1947, *Lecture Series in Nuclear Physics* (MDDC 1175) (Washington: U.S. Government); 1951, *Science*, **113**, 101.

Weisskopf, V. F., and Ewing, D. H., 1940, *Phys. Rev.*, **57**, 472.

Wilks, J., 1957, *Rep. Progr. Phys.*, **20**, 38, (London: Physical Society).

Wilson, A. H., 1953, *Theory of Metals* (Cambridge: University Press).

Wohlfarth, E. P., 1953, *Rev. Mod. Phys.*, **25**, 211.

Wolff, P. A., 1953, *Phys. Rev.*, **92**, 18.

Wonssowski, S. W., 1954, *Fortschr. Phys.*, **1**, 239.

Yang, C. N., and Lee, T. D., 1952, *Phys. Rev.*, **87**, 404.

Yevick, G. J., and Percus, J. K., 1956, *Phys. Rev.*, **101**, 1186.

Yvon, J., 1934, *Actualités Sci.*, No. 203.

Zener, C., and Heikes, R. R., 1953, *Rev. Mod. Phys.*, **25**, 191.

Ziman, J. M., 1952 a, b, *Proc. Phys. Soc.* A, **65**, 540, 548; 1953 a, *Ibid.*, **66**, 89; 1953 b, *Proc. Roy. Soc.* A, **219**, 257.

Zubarev, D., 1953, *J. Exp. Theor. Phys. USSR*, **25**, 548.

K MESONS AND HYPERONS
THEIR STRONG AND WEAK INTERACTIONS

By R. H. DALITZ

The Enrico Fermi Institute of Nuclear Studies, University of Chicago

CONTENTS

PAGE

§ 1. Introduction 164
§ 2. Theoretical proposals concerning the strange particles 170
    2.1. Comparison of production and decay processes 170
    2.2. Charge independence and isotopic spin 171
    2.3. The Gell-Mann–Nishijima isotopic spin assignments 175
    2.4. The strangeness number 178
    2.5. Strangeness and isotopic spin 180
§ 3. Laboratory production of strange particles 182
    3.1. Production processes in pion–hydrogen collisions 182
    3.2. Gamma ray observations on strange particle production 188
    3.3. The production of charged K particle beams 190
§ 4. The charged K particles and their interactions 192
    4.1. Mass measurements for charged K particles 192
    4.2. The two-body K particle decays 193
        4.2.1. The $K_{\mu 2}$ particle 193
        4.2.2. The $K_{\pi 2}$ particle 194
    4.3. The τ particle 195
    4.4. Further three-body K particle decays 206
        4.4.1. The $K_{\pi 3}$ or τ' decay 206
        4.4.2. The $K_{\mu 3}$ particle 208
        4.4.3. The K_{e3} particle 209
    4.5. Nuclear interactions of K^+ particles 210
    4.6. The relative frequencies of K^+ decay modes 216
    4.7. The nuclear interaction and decay of K^- particles 218
§ 5. The neutral K particles 228
    5.1. The θ^0 decay 228
    5.2. The relationship between K^0 and $\overline{K}^0$ particles 229
    5.3. Other K^0 particles and K^0 decay modes 235
        5.3.1. The τ^0 particle 236
        5.3.2. K^0 decays with identified secondaries 238
        5.3.3. Long-lived K^0 decays 239
        5.3.4. The time dependence of 'anomalous' θ^0 and τ^0 decay 241
§ 6. The decay and nuclear interactions of hyperons 243
    6.1. The decay modes of Λ and Σ hyperons 243
    6.2. The nuclear interactions of the Λ^0 hyperon 246
    6.3. The nuclear interactions of the Σ hyperons 258
    6.4. The Ξ hyperons 261
§ 7. Relationships between the K particle decay modes 262
    7.1. Electromagnetic transitions between K particles ? 263
    7.2. Non-conservation of parity in weak decays ? 265
    7.3. The hypothesis of invariance under parity conjugation 269
    7.4. Schwinger's model for weak decays 275
§ 8. Anomalous events 278
    8.1. Unstable nuclear fragments with anomalous Q values 278
    8.2. Events with K particle secondaries 279

PAGE
§ 9. Conclusion 281
9.1. The Gell-Mann–Nishijima scheme and particular theoretical models 281
9.2. Outlook 286
Acknowledgments 289
Appendix 289
References 298

Abstract. This article surveys our empirical knowledge of the production processes, the nuclear interactions and the decay processes of K particles and hyperons, as this appeared at the close of 1956, and interprets these phenomena within the framework of the charge-independent classification scheme proposed by Gell-Mann and by Nishijima for these ‘ strange particles ’. The assignment of a ‘ strangeness ’ quantum number to each strongly interacting particle is a deep-lying aspect of this scheme, and the main qualitative features of the empirical data allow a direct interpretation in terms of a law of *conservation of strangeness* through all strong interactions. Where appropriate, more detailed analyses of the phenomena are made on the assumption that charge independence holds for these strong interactions.

The evidence on decay processes for K particles and hyperons is discussed in some detail. These processes violate the strangeness conservation law and also involve changes in the total isotopic spin for those cases where this quantity is relevant for the final particles. The success of Gell-Mann's $\Delta T = \frac{1}{2}$ rule in accounting for certain aspects of the decay data is considered for each of the latter cases ; this rule is given a simple origin in Schwinger's model for strange particle decay processes, which is discussed briefly. The implications of symmetry between particle and antiparticle, as pointed out by Gell-Mann and Pais for the neutral K particles and their decay modes, is also discussed in some detail, together with an account of the experimental data on long-lived K^0 particles and anomalous K^0 decay modes.

Indications in the data which bear on the spin and parity values for K particles and hyperons are stressed, especially for the decay modes associated with τ and θ particles. Although it is shown that the products of τ and θ decay necessarily differ in spin or in parity, a great similarity has been found concerning all other properties of the τ and θ particles, and the proposals which have been made to account for this situation are surveyed. For the Λ^0 particle, observations on bound Λ^0 hypernuclei provide much information relevant to the Λ^0–nucleon interaction and the Λ^0 spin and parity, and this is also discussed in some detail.

A section is devoted to discussion of ‘ anomalous events ’ for which no interpretation is known in terms of the interactions of the well-established particles. These events and the possibility that heavier unstable particles or perhaps resonance states may exist have an obvious importance, of special relevance for attempts to construct a theoretical foundation leading naturally to the Gell-Mann–Nishijima scheme. A brief discussion of the preliminary attempts to establish such a foundation is given.

An appendix has been added in proof to discuss the recent observations on the failure of parity conservation and of charge conjugation invariance in certain weak decay processes, and the effects which a failure of these symmetry properties also for strange particle decays would have on the arguments made and the conclusions reached in this survey.

§ 1. INTRODUCTION†

In 1947, Lattes, Occhialini and Powell first established the existence in cosmic radiation of an unstable particle with a strong nuclear interaction. This particle, named the π meson or ‘ pion ’, was found to be copiously produced in the atmosphere by the impact of the incident cosmic-ray particles—its relative absence in cosmic-ray experiments at sea-level was explained by its instability in the free state, with a lifetime $2{\cdot}55(\pm 0{\cdot}1)\times 10^{-8}$ sec for decay to the lighter weakly interacting μ meson (or muon) and a neutrino. This particle, the pion, was then identified with the 'nuclear force meson' postulated by Yukawa in 1935. The nuclear forces between neutrons and protons were considered by Yukawa to result from the mutual exchange of the massive quanta of an intermediary meson field strongly coupled to neutrons and protons, and consequently to have a range of the order of $\hbar/m_\pi c$ where m_π denotes the mass of these meson quanta. The mass of the strongly interacting pion ($273{\cdot}0(\pm 0{\cdot}5)\ m_e$ (electron masses)) was in fact about that needed to account for the observed ranges of the nuclear forces.

---

† In this article it will normally be assumed that the units are chosen such that the velocity of light c and Planck's constant $\hbar$ each have the value unity. However these symbols will sometimes be retained in the formulae in order to display the dimensional character of the quantities considered.

After laboratory production of the pion was achieved in 1948 by the collision of cyclotron-accelerated alpha particles with various targets, pion beams became available for many detailed experiments on the properties of the pion. As a result, measurements became possible for the reactions responsible for the absorption of pions in hydrogen and deuterium, and these indicated that the pion was a *pseudoscalar* particle, i.e. that the pion was a spinless particle carrying odd intrinsic parity. Further experiments on the scattering of pions by hydrogen and on their production from hydrogen by high-energy electromagnetic radiation have established that the interaction of pions with nucleons does have many of the features needed to account for the detailed nature of nuclear forces and for other properties of the nucleons, following the general proposals of Yukawa. Our present knowledge on the properties and interactions of the pion has recently been ably summarized by Gell-Mann and Watson (1954) and by Bethe and de Hoffmann (1955), and we shall not discuss these in further detail here.

Even during this period of intense investigation of the pion properties, however, there was clear evidence for the existence of still heavier unstable particles amongst the cosmic-ray particles. During a cloud chamber investigation of penetrating showers carried out at Manchester, Rochester and Butler (1947) found an event which represented the spontaneous decay of a neutral unstable particle, of mass at least $800m_e$, into two singly charged particles. This discovery was soon confirmed by other cloud chamber groups, especially in work carried out at mountain altitudes, and as the experimental techniques became more refined, it was realized that these heavier short-lived particles occurred with a frequency of about several per cent relative to the pions. Such copious production suggested that these new unstable particles also had strong nuclear interactions and should also play a significant role in the theoretical interpretation of nuclear properties.

By 1951, the Manchester group (Armenteros, Barker, Butler and Cachon 1951) were able to show that there existed at least two classes of these 'neutral V particles', as they were then termed. In particular, they established that one of these particles decayed to a proton and a π^- meson with an energy release of about 40 MEV kinetic energy in its initial rest system; this particle has since been named the Λ^0 particle,

$$\Lambda^0 \to P + \pi^- + 36{\cdot}9(\pm 0{\cdot}2)\ \text{MEV}. \qquad \ldots\ldots(1.1)$$

More recently Λ^0 decay events have been identified in emulsion studies and the Q value (kinetic energy release) quoted in equation (1.1) was obtained from emulsion work reported by Friedlander, Keefe, Menon and Merlin (1954). The Λ^0 lifetime has been shown to be close to 3×10^{-10} sec (cf. § 6.1) from the observed spatial distribution of Λ^0 decay events in cloud chambers, following the method of estimation discussed by Bartlett (1953). A particle of this type, with a decay which leads finally to a neutron or proton, will now be referred to as a supernucleon or *hyperon.*

Many other neutral V events were reported in this initial period, but it was not until the painstaking work of Thompson, Buskirk, Etter, Karzmark and Rediker (1953) that the existence of another definite two-particle decay mode was clearly established; this second neutral particle is now termed the θ^0 particle, with the decay mode

$$\theta^0 \to \pi^+ + \pi^- + 214(\pm 5)\ \text{Mev.} \qquad \ldots\ldots(1.2)$$

Its lifetime was shown by later cloud chamber work to be about half that of the Λ^0 particle (cf. § 5.1). From the energy released in this decay, it follows that the particle has mass $966(\pm 10)\,m_e$. Such a particle, whose mass lies between those of the pion and of the proton, will be referred to as a *K particle.*

In this investigation of Thompson *et al.* it was still clear that other less frequent neutral V decays remained to be identified. Some recent events of this type will be described in § 5.3.

The existence of unstable charged K particles was established rather early and with particular clarity in the use of electron-sensitive nuclear emulsions exposed to cosmic radiation. In 1949 the Bristol group (Brown, Camerini, Fowler, Muirhead, Powell and Ritson 1949) described a particle which came to rest in emulsion and decayed to give three charged pions. That these were the only decay products was clear from the coplanarity of the three pion tracks and from the fact that their momenta had vector sum zero, within the errors of measurement. This particle was named the τ^+ meson and its decay mode was thus established to be

$$\tau^+ \to \pi^+ + \pi^+ + \pi^- + 75{\cdot}0(\pm 1{\cdot}5)\ \text{Mev.} \qquad \ldots\ldots(1.3)$$

With confirmation of this decay scheme from other groups, the τ^+ meson became the first well-established K particle and, already in this early work, it became clear (Herz, Hodgson and Tennent 1953) that the lifetime of the τ^+ particle was relatively long, exceeding 10^{-9} sec. The energy release quoted in equation (1.3) is that given by Amaldi (1956) on the basis of 54 τ^+ decays in stripped emulsions exposed to cosmic radiation; it corresponds to a mass of $966(\pm 4)\ m_e$ for the τ meson, rather close to that of the neutral K particle, the θ^0 meson.

Other charged K particles also became established. During an investigation of the electron spectrum in μ decay, O'Ceallaigh (1951) found an example of a positive K particle whose decay gave a low-energy μ meson, identified by its decay after coming to rest. In the next few examples observed, quite different energies were found for the secondary μ mesons, and it was concluded that this was a three-body decay process—the detailed evidence on this decay scheme will be summarized in § 4.4. With our present terminology, this particle is referred to as the $K^+_{\mu 3}$. This notation indicates that its decay leads to three final particles, of which one is a positively charged μ meson. In a similar way the $K^+_{\pi 3}$ decay, first reported by Crussard, Kaplon, Klarmann and Noon (1954), consists of a K particle decay leading to a positive pion of non-unique energy.

The identification of K particle decays leading to a charged secondary of relativistic velocity was more difficult in the early work, since the long range of the secondary made it very probable that the secondary would escape from the emulsion and evade identification. The presence of a $K^+_{\mu 2}$ particle was first established in work of the Paris group (Gregory, Lagarrigue, Leprince-Ringuet, Muller and Peyrou 1954) using a double cloud chamber; similarly, the first clear indication of the $K^+_{\pi 2}$ particle was a remarkable cloud chamber event reported by the Princeton group (Hodson, Ballam, Arnold, Harris, Rau, Reynolds and Treiman 1954). Later, when the use of large emulsion stacks became general, more accurate measurements on these particles became possible (G-stack collaboration 1956) and a further particle, the K^+_{e3} particle, was also identified (Friedlander, Fujimoto, Keefe and Menon 1955).

Although all of these K mesons were clearly demonstrated first in cosmic-ray work, laboratory experiments involving high-energy accelerators have given our most extensive information concerning their properties. This will be discussed in some detail in §§ 3, 4 and 5. One of the major questions which this information raises is the possibility of relationships amongst these many K particles, which have been named in a manner descriptive only of the decay process by which they are recognized, whether they may not represent alternate decay modes of a smaller group of particles or even of one K particle. This problem depends very much on the properties discussed in §§ 4 and 5, and on their analysis; this question will be raised in some detail in § 7.

The presence of charged V decays was also found very early in cloud chamber studies. Both positive and negative V decays were found, but with a frequency of only 10–20% relative to the neutral V decays. Many of them represent the decay of the K particles mentioned above, or of their negative counterparts; their relative fewness is related to the fact that a relativistic particle will cross a typical chamber in a time short compared with the K particle lifetimes. However, this evidence was initially important in establishing the existence of negative K particles corresponding to at least the most prominent of the positive K particles. In emulsions, negative K particles will generally come to rest before decay, undergo capture into Bohr orbits within the atoms constituting the emulsion and cascade down through successive atomic orbits until they reach a low-lying level and are absorbed by interaction with the atomic nucleus. Although the existence of K^- particles was clear in emulsion work, their relationship with the positive K particles was difficult to establish there. Only quite recently has the scanning of metres of K track for decays-in-flight in emulsion become feasible, with exposure to artificially produced K particles (cf. § 3.3 and § 4.7).

Positive and negative hyperons of lifetime about 10^{-10} sec also constitute a large proportion of these charged V decays. One of these hyperons is the 'cascade particle', whose decay mode

$$\Xi^- \to \Lambda^0 + \pi^- + 65(\pm 4) \text{ MeV} \qquad \ldots\ldots(1.4)$$

was first established clearly by Cowan (1954). Its characteristic appearance is of a charged V decay, giving a negative pion, followed by a Λ^0 decay from the neutral secondary, giving rise to the descriptive title 'cascade' hyperon. About six examples of this particle, each of them negatively charged, have been recorded to date, with *Q* values ranging between 60 and 70 MeV; no evidence exists to suggest a positive counterpart to this hyperon. At the present time the Ξ^- particle has not been reported in accelerator work; its rarity may be associated with a high threshold energy for its production (see § 2.4 and § 6.4).

The remaining well-established hyperons are the Σ particles. The Σ^+ particle has two closely related decay modes

$$\Sigma^+ \begin{cases} \rightarrow P + \pi^0 + 116{\cdot}1(\pm 0{\cdot}5) \text{ MeV} & \ldots\ldots(1.5a) \\ \rightarrow N + \pi^+ + 110(\pm 1) \text{ MeV.} & \ldots\ldots(1.5b) \end{cases}$$

The presence of a positive hyperon decay leading to a proton was first indicated in the cloud chamber work of the Pasadena group (York, Leighton and Bjornerund 1953). After the first identification of this Σ^+ decay in emulsion by the Milan group (Bonetti, Levi-Setti, Panetti and Tomasini, 1953), the unique proton range found for the emulsion examples of Σ^+ decay from rest established that the Σ^+ decay was a two-body process. From this proton range the *Q* value for the decay (1.5*a*) can be determined rather accurately; the value quoted here is that of Fry, Schneps, Snow and Swami (1956) based on twelve Σ^+ events. The alternative decay mode was also reported first by the Milan group (Bonetti, Levi-Setti, Panetti and Tomasini 1953) and examples are now available where the positive pion secondary is brought to rest and clearly identified. The nature of the neutral particles in these decays (1.5) has been deduced from the mass measurement for the primary Σ^+ and from the measured momentum for the two-body decay; the assignment adopted is supported by the close relationship between the two decay modes corresponding to the well-known charge-exchange reaction between pion and nucleon. The *Q* value quoted for (1.5*b*) has been obtained from the former *Q* value on the basis of this assumption; it is in close accord with the less accurate values obtained experimentally assuming the neutral secondary in (1.5*b*) to be a neutron.

The first clear indication of a related Σ^- decay appeared in diffusion cloud chamber work with the pion beam at Brookhaven (cf. § 3.1) and was reported by Fowler, Shutt, Thorndike and Whittemore (1954). In emulsions the Σ^- particle proved rather elusive at first since its decay can be observed only in flight, the Σ^- particle undergoing nuclear absorption after coming to rest in emulsion. However, a fortunate event recorded by Debenedetti, Garelli, Tallone and Vigone (1954*b*) in which the π^- secondary was emitted backwards from decay in flight and was sufficiently slow to come to rest, provided a firm identification of the charged secondary. The mass estimates for the primary Σ^- particle in these events indicate the decay process

$$\Sigma^- \rightarrow N + \pi^- + 118(\pm 2{\cdot}6) \text{ MeV,} \qquad \ldots\ldots(1.6)$$

where the *Q* value adopted is that given recently by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956). From this determination it appears that the Σ^- particle is about 16 electron masses heavier than the Σ^+ particle.

These particles are the K mesons and hyperons with which we shall primarily be concerned in this report. Their known properties will be discussed in more detail in §§ 4, 5 and 6. Apart from the introductory remarks above, there will be no attempt here to trace out the complicated sequence of experiments and discussion which developed during the period over which these particles were first detected and finally recognized as well established. Several years ago, as described for example in the previous report (Rochester and Butler 1953) on new unstable particles in this publication, the number of decay schemes needed to account for the observations appeared to be increasing rapidly. However, with the exception of the Σ particles, the particles now well established had all been proposed by that time and it is now clear, as a result of the development of more powerful cloud chamber techniques and the widespread use of large emulsion stacks, that many of the other decay processes considered at that time are very rare, if at all significant. There has therefore been a great clarification in the problems which are now before us. Now that the predominant particles are well established, the determination of their intrinsic parameters, their spins, parities, etc., and of their interactions and inter-relationships, have become our main concern. Clearly these parameters must be known before any well-founded steps can be taken toward the basic question posed by these particles, to understand the reason for matter to occur in such a variety of discrete forms.

At this point it should be emphasized that a number of outstanding events have been reported for which no simple interpretation in terms of the above particles seems to be available. The clearest of these events will be described and discussed in § 8, as it is important to emphasize the possibility that our present list of particles may still be far from complete. This is especially the case now, when the experimental programmes are concentrated in an energy range which may not be sufficiently high for appreciable production of any still heavier particles which may exist; the fact that the Ξ^- particle has not yet been reported in accelerator experiments has already been noted above.

Our empirical knowledge of the fundamental properties and interactions of these *strange particles* (as the K particles and hyperons will be generally termed) has increased with especial rapidity over the past two years. Proton accelerators giving proton beams with energies of some thousands of MEV have come into operation at several centres and have made possible the laboratory production of K mesons and hyperons in appreciable intensities, and consequently also laboratory experiments concerning the fundamental interactions of these particles in simple situations under well-defined conditions. It is with this quantitative information and its interpretation that we shall primarily be concerned in this report. We will find that, at the present time, our knowledge of these fundamental processes and properties is really very limited and uncertain in many respects. On the other hand, experiments especially informative on the elementary interactions are now becoming possible with the development of the hydrogen bubble chamber. At the present time, only preliminary results (cf. § 4.7) on the interaction of K^- particles with protons have been reported, but these are sufficient to indicate the power of this new technique. Very significant developments in this field may consequently be expected by the time this report appears; for this reason the current data will not be discussed in exhaustive detail but emphasis will be given to those aspects of the data which clearly illustrate various significant properties of the elementary interactions, or which indicate promising lines of future development.

Our interpretation and understanding of many of the experimental results now available has been influenced greatly by the initial successes of a classification scheme for strange particles proposed by Gell-Mann (1953) and independently by Nakano and Nishijima (1953). This scheme led naturally to many significant questions which have stimulated corresponding experimental research. To a considerable degree, the major progress in this field to the present has consisted in the verification of a wide variety of the qualitative predictions of the theory. It is therefore natural that our first task will be to discuss in § 2 the theoretical considerations which led to these theoretical proposals, and the nature of the predictions possible from them.

§ 2. THEORETICAL PROPOSALS CONCERNING THE STRANGE PARTICLES

2.1. *Comparison of Production and Decay Processes*

The interaction of K particles and hyperons with pions and nucleons must be strong interactions, comparable with the pion–nucleon interaction, in order to account for the cross sections (several per cent of geometrical) observed for strange-particle production in high-energy nuclear collisions. On the other hand, the decays of strange particles to systems of pions and nucleons have lifetimes of order 10^{-10} sec or longer, extremely long on the nuclear scale. For example, if the effective coupling parameter for the decay $\Lambda^0 \rightarrow P + \pi^-$ be denoted by g, then comparison of the observed Λ^0 lifetime with the standard expression (cf. Fermi 1951) for the lifetime of a two-body decay leads to the value $g^2/\hbar c = (4{\cdot}7 \times 10^{-6})^2 e^2/\hbar c$, a coupling constant many orders of magnitude smaller than that of the electromagnetic interaction. Clearly a single $(\Lambda^0 P \pi^-)$ interaction could account both for the strong Λ^0 production and the weak decay processes only if it had a very strong momentum dependence, so that it was very weak for the low momenta allowed by the small energy release in the decay and very strong for the high momenta which are available in high-energy collisions. Fortunately this possibility is clearly excluded for the Λ^0 particle by the observation that its lifetime is not strongly modified by the presence of neighbouring nucleons (cf. § 6.2), although pions of high momentum are available in their pion clouds. Other unattractive features of such a proposal to account for the observed situation have been emphasized also by Gell-Mann and Pais (1954). Consequently this great contrast between the strengths of the production and decay interactions indicates that these must be quite distinct interactions and excludes the possibility that single production of strange particles should be appreciable in collisions of nucleons and pions.

From this conclusion, Pais (1952) was led to the hypothesis of *associated production* for strange particles. According to this hypothesis the strong production interactions are always such that at least two strange particles are involved in the process, whilst interactions concerned with only one strange particle, as in the decay process, are extremely weak. This hypothesis is supported by the strange-particle production events which have been identified in collisions of pions and protons in hydrogen (see § 3.1); for example, there are now many examples of the process

$$\pi^- + \mathrm{P} \rightarrow \Lambda^0 + \theta^0, \qquad \ldots\ldots(2.1)$$

where both Λ^0 and θ^0 particles are identified by their subsequent decay. Although not all strange-particle production events display associated production in this explicit way (owing, for example, to escape of Λ^0 or θ^0 particle before decay), there have been no cases reported which are inconsistent with interpretation as associated-production events. These associated-production interactions do not conflict with slow decay for strange particles since they can only transform one strange particle into a system involving another strange particle. For example, although the strong interaction $\Lambda^0 \rightleftharpoons N + \bar{\theta}^0$ implied by equation (2.1) ($\bar{\theta}^0$ denotes the antiparticle to the θ^0 particle) can effect a transition from Λ^0 particle to a virtual $N + \bar{\theta}^0$ state, it does not give rise to a rapid decay process for the Λ^0 particle as the Λ^0 mass does not provide sufficient energy for this to be a real process.

One may now look for selection rules by means of which the strong ' associated ' interactions are distinguished from the weak decay interactions. There are, of course, already some selection rules effective in these processes. Clearly *conservation of charge* is necessary for all interactions. Also the very existence of nuclei requires that there should exist no processes involving destruction (or creation) of a nucleon without corresponding annihilation (or creation) of an anti-nucleon; since the hyperons lead ultimately to a proton or neutron, this statement must be extended to include both nucleons and hyperons. This extension is referred to as *conservation of baryons*—' baryon ' being a generic term for nucleon or hyperon—and states that the total number of nucleons and hyperons less the total number of anti-nucleons and anti-hyperons must remain unchanged through any reaction. Naturally these conservation laws apply to all interactions, both strong and weak.

In this search, it is instructive to look first at the properties of the particle interactions with which we already have some familiarity. These may be distinguished clearly as of three kinds: (i) strong nuclear interactions, such as the pion–nucleon and nucleon–nucleon interactions, (ii) electromagnetic interactions, which may all be traced back (through the principle of gauge invariance) to the electromagnetic interaction with the charge of each particle, (iii) weak interactions, such as the beta-interaction responsible for neutron decay and muon decay.

For muons and electrons, the processes observed are satisfactorily accounted for by interactions of classes (ii) and (iii). Strange particles apart, the only strong interactions known are those between pions and nucleons; the outstanding property common to these interactions is their *charge independence*. The most obvious manifestation of this property (to be discussed in the next section) is the appearance of the strongly interacting particles as two multiplets of particles, the three pions and the two nucleons, the members of each multiplet having different charges but essentially the same mass.

Among the strange particles, it is clear that similar multiplets exist; the K^+, K^0, K^- particles have essentially equal masses, as is also the case for Σ^+, Σ^- particles. This suggests that some form of charge independence may exist for strange particles and their strong interactions also. For this reason, the nature of charge independence for the pions and nucleons will now be discussed in some detail.

2.2. *Charge Independence and Isotopic Spin*

The principle of charge independence of nuclear forces states that, for a given energy of relative motion, the N–N, N–P and P–P nuclear forces depend only on the total angular momentum and parity of the two particles and not on their charges. This has been established rather precisely for the N–P and P–P nuclear forces in the singlet S states at low energies. The equality of N–N and P–P forces in a given state is already known as the weaker principle of *charge symmetry*, for which the strongest evidence comes from the close correspondence between the levels of two mirror nuclei (after correction for the difference in Coulomb effects).

Cassen and Condon (1936) first pointed out that this principle could be very simply expressed by use of the concept of *isotopic spin* for the nucleons. The proton and neutron are regarded as the two substates of nucleon doublet of isotopic spin $T = \frac{1}{2}$, the isotopic spin component T_3 having value $+\frac{1}{2}$ for the proton, $-\frac{1}{2}$ for the neutron. If u_P, u_N denote isotopic spin wave functions for the proton state and the neutron state respectively, then

$$T_3 u_P = +\tfrac{1}{2}u_P, \quad T_3 u_N = -\tfrac{1}{2}u_N.$$

The charge Q of the nucleon is then related to its isotopic spin value T_3 by the relation

$$Q = T_3 + \tfrac{1}{2}. \qquad \ldots\ldots(2.2)$$

For a system of two nucleons the total isotopic spin may have values $T = 1$ or $T = 0$, following a vector addition of two isotopic spins of magnitude $\frac{1}{2}$. For two protons, the total component T_3 is $+1$ and the isotopic spin wave function is $u_P(1)\,u_P(2)$; similarly, for two neutrons, the total T_3 is -1, the isotopic spin wave function is $u_N(1)\,u_N(2)$. The PP and NN systems therefore clearly have total isotopic spin $T = 1$; their isotopic spin states $u_P(1)\,u_P(2)$ and $u_N(1)\,u_N(2)$ are symmetrical between the two particles. For the NP system, where the total $T_3 = 0$, two isotopic spin states may be written down,

$$[u_P(1)\,u_N(2) + u_N(1)\,u_P(2)]/\sqrt{2} \quad \text{and} \quad [u_P(1)\,u_N(2) - u_N(1)\,u_P(2)]/\sqrt{2},$$

of which the former is symmetrical between the two particles, the latter antisymmetrical. It is natural to associate the former with the symmetrical PP and NN states and therefore with total $T = 1$, whereas the latter corresponds to total $T = 0$. These wave functions may be arrayed as a $T = 1$ charge triplet and a $T = 0$ charge singlet:

$$\left.\begin{array}{l} \left.\begin{array}{l} \phi_{1,+1} = u_P(1)\,u_P(2), \\ \phi_{1,0} \;\;= [u_P(1)\,u_N(2) + u_N(1)\,u_P(2)]/\sqrt{2}, \\ \phi_{1,-1} = u_N(1)\,u_N(2), \end{array}\right\} T = 1, \\ \phi_{0,0} \;\;= [u_P(1)\,u_N(2) - u_N(1)\,u_P(2)]/\sqrt{2}, \quad T = 0. \end{array}\right\} \qquad \ldots\ldots(2.3)$$

In these equations (and below), the notation $\phi_{T,m}$ denotes the isotopic-spin wave function for a state of total isotopic spin T and isotopic spin component $T_3 = m$. These expressions (2.3) are identical in form with the wave functions for the triplet and singlet states of ordinary spin, obtained by combining the spins of two spin-$\frac{1}{2}$ particles, although isotopic spin and ordinary spin are physically unrelated quantities.

This provides a special example of the general result that the isotopic spins of a group of particles may be added to give various values of total isotopic spin, following the same vector addition rules as those applicable to addition of angular momenta (cf. Blatt and Weisskopf 1952, Rose 1955).

According to the Pauli principle, the PP and NN states must be antisymmetrical in space and spin variables: with a symmetrical isotopic spin wave function, their total wave function is antisymmetrical in space, spin and isotopic spin variables. This suggests an extended Pauli principle for nucleons, requiring a wave function totally antisymmetric in space, spin and isotopic spin variables. Applying this principle, the $T = 1$ NP state (with symmetric isotopic spin wave function) must have an antisymmetric space and spin wave function: the nuclear interaction effective in this state must then be the same as that in the corresponding PP and NN states, according to the charge independence principle. The $T = 0$ NP state, having an antisymmetric isotopic spin wave function, must have a wave function symmetric in space and spin variables; clearly this state has no counterpart in the PP and NN systems. With this extended Pauli principle, the statement of charge independence is that the nuclear interaction depends only on the total isotopic spin T of the state considered, and not on the component T_3 which specifies the total nuclear charge in this state.

For light nuclei, where the charge-dependent Coulomb forces play a minor role, charge independence requires each nuclear energy level to correspond to a definite total isotopic spin T. For this system, the total charge will be given by

$$Q = T_3 + \tfrac{1}{2}B, \qquad \ldots\ldots(2.4)$$

where B is the total number of nucleons. The state T will have $2T+1$ substates with values of T_3 ranging from $-T$ to $+T$, so that sets of $2T+1$ isobaric nuclei of mass number B are expected, each with an energy level of the same binding energy (after allowance for the Coulomb energies). This expectation is very well confirmed in light nuclei, for example in the isobaric triad ^{10}Be, ^{10}B, ^{10}C. The evidence on charge independence from nuclear energy levels and from nuclear reactions in light nuclei has been summarized by Wilkinson (1953) and by Burcham (1955), and provides very strong support for the validity of isotopic spin conservation.

Charge independence for the pion–nucleon interaction was first proposed by Kemmer (1938), who recognized that this would provide a direct interpretation of charge independence for nuclear forces. Kemmer considered the π mesons as the three substates of a $T = 1$ isotopic spin state, the π^+, π^0, π^- mesons being the substates with $T_3 = +1, 0, -1$ respectively. For the pion, therefore,

$$Q = T_3. \qquad \ldots\ldots(2.5)$$

This is a special case of (2.4), since there are no nucleons involved here and therefore $B = 0$. The interaction of a pion and nucleon with definite total angular momentum j and parity w then depends only on their total isotopic spin T, which may have the values $T = \frac{3}{2}$ or $T = \frac{1}{2}$. The system $\pi^+ + \mathrm{P}$ is typical of $T = \frac{3}{2}$, since these particles have a total T_3 of $+\frac{3}{2}$. However, the systems $\pi^- + \mathrm{P}$, $\pi^0 + \mathrm{N}$ which have a total T_3 of $+\frac{1}{2}$ are mixtures of $T = \frac{3}{2}$ and $T = \frac{1}{2}$ states. If we denote the π meson charge states by $\mathrm{U}_+, \mathrm{U}_0, \mathrm{U}_-$, the combinations of these pion–nucleon systems which have definite total isotopic spins are as follows :

$$\left.\begin{array}{l}\left.\begin{array}{l}\phi_{\frac{3}{2},+\frac{3}{2}} = u_{\mathrm{P}}\,\mathrm{U}_{+},\\ \phi_{\frac{3}{2},+\frac{1}{2}} = \sqrt{\tfrac{2}{3}}u_{\mathrm{P}}\,\mathrm{U}_{0}-\sqrt{\tfrac{1}{3}}u_{\mathrm{N}}\,\mathrm{U}_{+},\\ \phi_{\frac{3}{2},-\frac{1}{2}} = \sqrt{\tfrac{1}{3}}u_{\mathrm{P}}\,\mathrm{U}_{-}+\sqrt{\tfrac{2}{3}}u_{\mathrm{N}}\,\mathrm{U}_{0},\\ \phi_{\frac{3}{2},-\frac{3}{2}} = u_{\mathrm{N}}\,\mathrm{U}_{-},\end{array}\right\} T=\tfrac{3}{2},\\ \left.\begin{array}{l}\phi_{\frac{1}{2},+\frac{1}{2}} = \sqrt{\tfrac{1}{3}}u_{\mathrm{P}}\,\mathrm{U}_{0}+\sqrt{\tfrac{2}{3}}u_{\mathrm{N}}\,\mathrm{U}_{+},\\ \phi_{\frac{1}{2},-\frac{1}{2}} = \sqrt{\tfrac{2}{3}}u_{\mathrm{P}}\,\mathrm{U}_{-}-\sqrt{\tfrac{1}{3}}u_{\mathrm{N}}\,\mathrm{U}_{0}.\end{array}\right\} T=\tfrac{1}{2}.\end{array}\right\} \quad \ldots\ldots(2.6)$$

From what has been said above, these functions $\phi_{T,m}$ clearly correspond to the $T_3 = m$ values stated. For $T_3 = \pm\frac{1}{2}$, the two pion–nucleon states possible have been combined with appropriate (Wigner) coefficients, chosen with the property that the final combinations given in (2.6) correspond to states of definite total isotopic spin T (cf. Bethe and de Hoffmann 1955).

For the pion–nucleon system, the interactions effective in the $T = \frac{3}{2}$ and $T = \frac{1}{2}$ states have been found experimentally to be quite different. For $T = \frac{3}{2}$, the pion–nucleon system has a strong resonance for a total kinetic energy (in centre-of-mass system) of the order of 160 MeV, whereas the interaction in the $T = \frac{1}{2}$ state is quite weak until relatively high energies.

With this isotopic spin formalism for the pions, Bose statistics holds for pions irrespective of their charge, so that the wave function for a system of particles must be unchanged when both space and isotopic spin variables of any two pions are interchanged. With $T = 1$ for each pion, the total isotopic spin of two pions may have the value $T = 0$, 1 or 2. Expressions for the corresponding isotopic spin states may be obtained by analogy with the standard formulae for the addition of two unit angular momenta (cf. Blatt and Weisskopf 1952, Rose 1955). For two pions whose total charge is zero ($T_3 = 0$), we will give these wave functions explicitly, so that they will be available for use later :

$$\left.\begin{array}{l}\phi_{0,0} = \dfrac{1}{\sqrt{3}}[\mathrm{U}_{+}(1)\,\mathrm{U}_{-}(2)-\mathrm{U}_{0}(1)\,\mathrm{U}_{0}(2)+\mathrm{U}_{-}(1)\,\mathrm{U}_{+}(2)],\\[2ex] \phi_{1,0} = \dfrac{1}{\sqrt{2}}[\mathrm{U}_{+}(1)\,\mathrm{U}_{-}(2)-\mathrm{U}_{-}(1)\,\mathrm{U}_{+}(2)],\\[2ex] \phi_{2,0} = \dfrac{1}{\sqrt{6}}\Big[\mathrm{U}_{+}(1)\,\mathrm{U}_{-}(2)+2\mathrm{U}_{0}(1)\,\mathrm{U}_{0}(2)+\mathrm{U}_{-}(1)\,\mathrm{U}_{+}(2)\Big].\end{array}\right\} \quad \ldots\ldots(2.7)$$

The functions $\phi_{0,0}$ and $\phi_{2,0}$ are symmetrical between the two particles 1 and 2 (this is characteristic also of the other $T = 2$ states $\phi_{2,\pm1}$, $\phi_{2,\pm2}$), whilst $\phi_{1,0}$ is antisymmetric for interchange of 1 and 2 (which is naturally characteristic also for the other $T = 1$ states, $\phi_{1,\pm1}$). Now, for two pions, the interchange of their spatial positions is identical with the parity operation on their relative motion. Hence Bose statistics for the two pions requires that states with $T = 0$ and 2 can occur only when their relative motion has even parity, states with $T = 1$ only with motion of odd parity.

Charge independence gives rise to many relationships amongst reactions involving pions and nucleons. For example, Hildebrand (1953) has compared the reactions (*a*) $P+P \rightarrow D+\pi^+$, (*b*) $N+P \rightarrow D+\pi^0$; since the deuteron has $T=0$, the final state has $T=1$. However, the system N+P is an equal mixture of states $T=0$ and $T=1$, whereas P+P is entirely $T=1$. Charge independence therefore requires that reaction (*a*) should have differential cross section just twice that for reaction (*b*), and the measurements of Hildebrand (1953) and of Schluter (1954) give a striking confirmation of this prediction. Similarly in the P+D reaction giving $^3H+\pi^+$ or $^3He+\pi^0$, the fact that $T=\frac{1}{2}$ for the initial state and therefore for the final state also, requires the π^+/π^0 ratio at all angles to be 2/1 (as is indicated by the $T=\frac{1}{2}$ wave function of equation (2.6)). This has been established by Bandtel, Frank, Madey and Moyer (1954).

The empirical evidence for the charge independence of the strong interactions of class (i) is therefore very substantial. Isotopic spin is not conserved in the electromagnetic interactions or the interactions of class (iii), and the violations of isotopic spin conservation which have been observed can be attributed to these. However, the electromagnetic interactions always leave the component T_3 of the system unchanged: for pions and nucleons this follows from the relation (2.4) and the fact that charge and baryon number are necessarily conserved separately. Since electrons and muons have no strong interactions and therefore no isotopic spin assignments, the known decay interactions (iii) always involve states for which no isotopic spin conservation properties can be expected.

2.3. *The Gell-Mann–Nishijima Isotopic Spin Assignments*

Charge independence for strange particles can be ensured most directly by assigning an isotopic spin to each strongly interacting particle and requiring that all strong interactions conserve isotopic spin. This was first carried out successfully by Gell-Mann (1953) and, independently, also by Nakano and Nishijima (1953). As before, the electromagnetic interactions will conserve the component of isotopic spin T_3, but will generally allow a change in the total isotopic spin. Any interactions which do not conserve total T or its component T_3 are assumed to be very weak, of class (iii).

Consider first the Λ^0 particle. Since this particle is a singlet state with no charged counterpart, it must correspond to $T=0$. From this it is clear that the interaction $\Lambda^0 \rightarrow P+\pi^-$ cannot be strong, since the isotopic spin T is necessarily half-integral in the final state; the total isotopic spin and its component ($T_3=-\frac{1}{2}$) each differ by a half-integer from the initial value, so that only an interaction of class (iii) is possible for this process. This example illustrates that, with this $T=0$ assignment for the Λ^0 particle, conservation of T_3 is no longer equivalent with conservation of charge, since the latter is satisfied in this interaction. Similarly, $\Lambda^0 \rightarrow N+$ photons involves a change from $T_3=0$ to $T_3=-\frac{1}{2}$ for the final neutron, and is therefore forbidden except through a weak decay interaction of class (iii).

On the same lines, a fast K^+ decay will be forbidden if isotopic spin $T=\frac{1}{2}$ is assigned to the K^+ particle; the observed decays to pion states then involve half-integral changes in isotopic spin component T_3 and can only be due to class (iii) interactions. The neutral K particle is then the second member of this $T=\frac{1}{2}$ doublet, for which

$$Q = T_3+\tfrac{1}{2}. \qquad \ldots\ldots(2.8)$$

As far as isotopic spin alone is concerned, all the various K^+ particles will be attributed to this $T = \frac{1}{2}$ doublet; this does not exclude the possibility that there may be several K particle states of differing angular momenta and parities included within this doublet.

With these assignments, the associated production $\pi^- + P \rightarrow \Lambda^0 + \theta^0$ may now proceed through a strong charge-independent interaction. The final state necessarily has total $T = \frac{1}{2}$, and $T_3 = -\frac{1}{2}$, whilst the initial state, which also has $T_3 = -\frac{1}{2}$, is partly in a $T = \frac{1}{2}$ and partly in a $T = \frac{3}{2}$ state. With the assumption of charge-independence this interaction can proceed only from the $T = \frac{1}{2}$ part of the initial state.

It should be noted here that not all conceivable associated-production processes are compatible with charge independence when these assignments are made. The reaction

$$N + N \rightarrow \Lambda^0 + \Lambda^0 \qquad \ldots\ldots(2.9)$$

is conceivable since it satisfies charge and baryon conservation. However, the final state has $T = 0$ whilst the initial NN state necessarily has $T = 1$. Now violations of total isotopic spin conservation are known in nuclear physics due to electromagnetic effects: however, the forbiddenness of this reaction (2.9) could not be relieved by electromagnetic effects since T_3 changes from a value of -1 initially to the value 0 finally. This example also makes it clear that T_3 conservation and charge conservation are no longer identical, and that the important distinction between allowed and forbidden reactions lies not so much in whether or not total T is conserved, but in whether or not the component T_3 is conserved.

Although the threshold for reaction (2.9) is quite low, about 0·77 GEV in the laboratory system, the evidence is that its occurrence is extremely rare, if it occurs at all (cf. § 3.2). Consequently the fact that these isotopic spin assignments do not allow this to be a strong charge-independent interaction is especially satisfactory.

The K^- particle may now be interpreted as the antiparticle to the K^+ particle. Since particle and antiparticle necessarily have opposite values of T_3, the K^- particle will have $T_3 = -\frac{1}{2}$. Its neutral counterpart with $T_3 = +\frac{1}{2}$ will then be the $\bar{K}^0$ particle, antiparticle to the K^0 particle. For these K particles (K^- and $\bar{K}^0$), the charge is related to the isotopic spin component by

$$Q = T_3 - \tfrac{1}{2}. \qquad \ldots\ldots(2.10)$$

Clearly a $\bar{K}$ particle can be created together with a K particle, this being the origin of the concept of antiparticle; a typical reaction would be

$$P + P \rightarrow P + P + K^+ + K^-. \qquad \ldots\ldots(2.11)$$

In fact these pair-creation reactions are the only production processes permitted for the $\bar{K}$ particle until extremely high energies are reached. For example, the conceivable reaction

$$\pi^- + P \rightarrow \Lambda^0 + \bar{K}^0 \qquad \ldots\ldots(2.12)$$

conserves total isotopic spin but does not conserve T_3, the final T_3 being $+\frac{1}{2}$ compared with an initial $T_3 = -\frac{1}{2}$. It is therefore forbidden except through the very weak interactions (iii), to which correspond entirely negligible cross sections.

Since the $\bar{K}^0$ particle has a different T_3 value from the K^0 particle, these are necessarily quite distinct entities. This situation provides a contrast with the case of the π^0 meson, which is its own antiparticle, and has some interesting consequences which will be discussed in § 5.2. That K^0 and $\bar{K}^0$ particles are distinct is given direct empirical support by the absence of the reaction (2.9). From (2.2) it is known that the interaction $N \rightleftharpoons \Lambda^0 + K^0$ is strong; if the K^0 particle were identical with its antiparticle, then it would follow that the interaction $\Lambda^0 \rightleftharpoons N + K^0$ would also be strong. From these the reaction

$$N + N \rightarrow \Lambda^0 + K^0 + N \rightarrow \Lambda^0 + \Lambda^0 \qquad \ldots\ldots(2.13)$$

would result as a strong process, contrary to rather detailed investigations (cf. § 3.2).

For the Σ particles, an integral isotopic spin assignment is necessary to ensure stability against a rapid decay to a pion and a nucleon. In the absence of doubly charged Σ particles, $T = 1$ is the natural choice, which gives the relationship

$$Q = T_3, \qquad \ldots\ldots(2.14)$$

the same as for the Λ^0 particle. Interactions of the type

$$\Sigma^{\pm} \rightleftharpoons \Lambda^0 + \pi^{\pm} \qquad \ldots\ldots(2.15)$$

then conserve both T and T_3 and can be charge-independent and strong; however, no fast decay can take place through this interaction since the $\Sigma - \Lambda$ mass difference is only of the order of 80 MEV. The interpretation of the Σ particles as a $T = 1$ multiplet implies the existence of a Σ^0 particle. For this particle, a fast radiative decay,

$$\Sigma^0 \rightarrow \Lambda^0 + \gamma, \qquad \ldots\ldots(2.16)$$

which conserves T_3, is allowed by the electromagnetic interactions. For this decay it would be natural to expect a lifetime of order 10^{-20} sec, longer than the natural time of $\hbar/m_\pi c^2$ because of the relative weakness of the electromagnetic coupling. The fact that no evidence exists for a slow pion decay $\Sigma^0 \rightarrow P + \pi^- + 110$ MEV was attributed by Gell-Mann to the competition of this rapid radiative decay (2.16). As a result of this radiative decay, which will take place very close to the point of origin of the Σ^0 particle, the existence of the Σ^0 particle must be established by indirect methods (cf. §§ 3.1, 4.7 and 6.3).

Σ particle production can occur jointly with K particles through strong interactions following reactions of the type

$$\begin{tikzcd}[row sep=small] & \Sigma^- + K^+, \quad \ldots\ldots(2.17a) \\ \pi^- + P \arrow[ur] \arrow[dr] & \\ & \Sigma^0 + K^0. \quad \ldots\ldots(2.17b) \end{tikzcd}$$

Since the K particles form only a doublet, it is clear that the Σ^+ particle cannot be produced in a $\pi^- + P$ collision near the threshold (0·90 GEV) for these processes; of course, the Σ^+ particle can still be produced at slightly higher energies with creation of a final pion,

$$\pi^- + P \rightarrow \Sigma^+ + K^0 + \pi^-. \qquad \ldots\ldots(2.18)$$

Nucleon–nucleon collisions leading to production of two hyperons alone are forbidden, as was the reaction (2.9); production of a hyperon and a K particle, for example,

$$\begin{tikzcd}[row sep=small] & \Sigma^0 + K^+ + P, \quad \ldots\ldots(2.19a) \\ P + P \arrow[ur] \arrow[dr] & \\ & \Sigma^+ + K^0 + P, \quad \ldots\ldots(2.19b) \end{tikzcd}$$

and closely related reactions, or production of two hyperons and two K particles if the energy is sufficiently high, can proceed through charge-independent interactions.

In contrast with equation (2.15), it will be of interest to remark that the assignment $T = 0$ for the Λ^0 particle implies

$$\Lambda^0 \not\leftrightarrow \Lambda^0 + \pi^0. \qquad \ldots\ldots(2.20)$$

This is forbidden as a strong interaction as the right-hand side has $T = 1$; it will be allowed as a rather weak interaction through virtual electromagnetic interactions. A strong interaction between a Λ^0 particle and a pair of pions, $\Lambda^0 \longleftrightarrow \Lambda^0 + \pi^+ + \pi^-$, can occur as a result of two operations of (2.15), $\Lambda^0 \longleftrightarrow \Sigma^+ + \pi^- \longleftrightarrow \Lambda^0 + \pi^+ + \pi^-$.

For the cascade Ξ particle, the situation is less clear. Only the Ξ^- particle has been observed experimentally; however, Gell-Mann postulated that a neutral Ξ^0 particle should also exist, forming the second member of a $T = \frac{1}{2}$ Ξ doublet, for which

$$Q = T_3 - \tfrac{1}{2}. \qquad \ldots\ldots(2.21)$$

The only decays related to the observed Ξ^- decay and allowed energetically for this Ξ^0 particle are $\Xi^0 \rightarrow \Lambda^0 + \pi^0$ or $\Xi^0 \rightarrow \Sigma^0 + \gamma$, neither of which are readily identifiable. The energetically most favourable production process which allows conservation of T and T_3 involves also the production of two K particles, for example

$$\mathrm{P} + \mathrm{P} \rightarrow \Xi^- + \mathrm{P} + \mathrm{K}^+ + \mathrm{K}^+, \qquad \ldots\ldots(2.22)$$

which has a threshold of 3·74 GeV.

$\Xi^-(T_3 = -\frac{1}{2})$	$\Xi^0(T_3 = +\frac{1}{2})$		$T = \frac{1}{2}, s = -2, U = -1.$	Hyperons
$\Sigma^-(T_3 = -1)$	$\Sigma^0(T_3 = 0)$	$\Sigma^+(T_3 = +1)$	$T = 1, s = -1, U = 0.$	
	$\Lambda^0(T_3 = 0)$		$T = 0, s = -1, U = 0$	
	$N(T_3 = -\frac{1}{2})$	$P(T_3 = +\frac{1}{2})$	$T = \frac{1}{2}, s = 0, \quad U = +1.$	Nucleons
	$\mathrm{K}^0(T_3 = -\frac{1}{2})$	$\mathrm{K}^+(T_3 = +\frac{1}{2})$	$T = \frac{1}{2}, s = +1, U = +1$	K particles
(————) $\mathrm{K}^-(T_3 = -\frac{1}{2})$	(————) $\overline{\mathrm{K}}^0(T_3 = +\frac{1}{2})$		$(T = \frac{1}{2}, s = -1, U = -1)$	($\overline{\mathrm{K}}$ particles)
$\pi^-(T_3 = -1)$	$\pi^0(T_3 = 0)$	$\pi^+(T_3 = +1)$	$T = 1, s = 0, \quad U = 0.$	Pions

Figure 1. The isotopic-spin assignments for the strongly-interacting particles in the Gell-Mann–Nishijima scheme. The strangeness number s and the 'hypercharge' $U = s + B$ are also tabulated.

2.4. *The Strangeness Number*

It has been emphasized above that the conservation of T_3 is the most important condition, for the charge-independence scheme, in distinguishing allowed from forbidden reactions. This is due to the variation in the linear relation between Q and T_3 from multiplet to multiplet. For strange particles, relation (2.4) no longer holds. Nakano and Nishijima (1953), Nishijima (1955) and Sachs (1955) have pointed out the advantages of using the parameters defined by the linear relation

$$Q = T_3 + \tfrac{1}{2}B + \tfrac{1}{2}s \qquad \ldots\ldots(2.23)$$

to characterize each multiplet. This parameter s has value 0 for pions and nucleons, $+1$ for K particles from equation (2.8), -1 for $\bar{K}$ particles (cf. equation (2.21)) and the Λ, Σ multiplets (cf. equation (2.15)), and -2 for the cascade doublet, according to equation (2.21). It will be termed the ***strangeness number*** for the multiplet.

With charge independence, and allowing for electromagnetic effects, reactions for which conservation of Q, T_3 and B are possible are the strong allowed reactions, all others occurring only through class (iii) interactions, if at all. Since conservation of Q and B is always necessary, a reaction is possible with strong interactions if and only if the total s is conserved, i.e.

$$\Delta s = 0 \textit{ for allowed reactions.} \qquad \ldots\ldots(2.24)$$

It should be emphasized here that these reactions are allowed only in so far as T_3 conservation is concerned; they may still be forbidden by other conservation laws (e.g. angular momentum and parity conservation) or perhaps weakened by failure to conserve total T (e.g. the interaction (2.20)). With use of s, discussion of the interactions possible becomes more transparent.

For example, since all hyperons have negative s, conservation of s clearly requires production of a K particle ($s = +1$) whenever a $\bar{K}$ particle ($s = -1$) is created in collisions of pions and nucleons; in nucleon–nucleon collisions the threshold (lab.) energy for $\bar{K}$ particle production is consequently 2·50 GEV. Similarly, production of two hyperons in a nucleon–nucleon collision cannot conserve s unless two K particles are also produced, a process with still higher threshold energy.

It is then clear that, with the absence of hyperons with positive s, a K particle interacting with a proton at moderate energy can only be elastically scattered. On the other hand, the $\bar{K}$ particle can undergo reactions with a nucleon to give a (Λ, Σ) hyperon with $s = -1$,

$$K^- + P \rightarrow \begin{cases} \Lambda^0 + \pi^0, \\ \Sigma^- + \pi^+, \\ \Sigma^0 + \pi^0, \\ \Sigma^+ + \pi^-, \end{cases} \qquad \ldots\ldots(2.25)$$

for example, in addition to the possibility of elastic scattering.

Similarly at low energies, the interaction of a Λ^0 particle with a nucleon can lead only to elastic scattering; at energies above approximately 150 MEV, reactions of the type $\Lambda^0 + P \rightarrow \Sigma^+ + N$ will also become possible. However, the Σ particle will interact rapidly with a neutron or proton at all energies, transforming to a Λ^0 particle with release of approximately 80 MEV kinetic energy; for example,

$$\Sigma^- + P \rightarrow \Lambda^0 + N. \qquad \ldots\ldots(2.26)$$

With $s = -2$ for the Ξ^- particle, its creation in a pion–nucleon or nucleon–nucleon collision needs simultaneous production of two K particles (since these are the only particles of positive s) for balance of total s. However, for incident K^- particles above 650 MEV, the Ξ^- particle can be produced through the reaction

$$K^- + P \rightarrow \Xi^- + K^+, \qquad \ldots\ldots(2.27)$$

although this endothermic reaction has to compete against the exothermic reactions (2.25). In the nuclear capture of the Ξ^- particle, its s value can be distributed between two Λ^0 particles,

$$\Xi^- + P \rightarrow \Lambda^0 + \Lambda^0, \qquad \ldots\ldots(2.28)$$

a process which would release about 30 MEV kinetic energy.

In all observed slow decay processes which lead to a final state to which a strangeness number may be assigned, there is observed a change in strangeness of one unit between initial and final state, so that

$$\Delta s = \pm 1, \textit{ for slow decay processes.} \qquad \ldots\ldots(2.29)$$

For the Ξ^- particle, the decay $\Xi^- \rightarrow N + \pi^-$, which is energetically much more favourable than the observed decay, has not been detected; this may be attributed to the fact that this 'cross-over transition' would involve a strangeness change $\Delta s = +2$.

2.5. *Strangeness and Isotopic Spin*

It will now be clear that these assignments of s and the selection rule (2.24) distinguishing strong interactions could be adopted whether or not isotopic spin were relevant for these particles. Strangeness s would then be an entirely new quantum number whose introduction would be guided by its success in classifying strong interactions. The assumption of charge independence is therefore not strictly essential to the general success of the scheme. However, it should be stressed here that the strangeness rules provide only a qualitative classification scheme for the possible interactions and do not lead to any detailed predictions of cross sections or of quantitative relationships between various processes; they provide only a framework for more detailed assumptions on spins, parities, etc., and on specific interaction terms.

It is the great merit of the Gell-Mann–Nishijima proposals that they lead to these selection rules for strange particles without introduction of any quantum numbers not previously known, simply by relaxation of the relationship between Q and T_3 so that T_3 conservation becomes independent from charge conservation. Also, without charge independence and corresponding isotopic spin assignments, the observed multiplet structure would not have a ready interpretation, although it is difficult to develop this argument for charge independence in a quantitative way. However, the presence of charge independence will give rise to many quantitative relationships between the various interactions connecting given multiplets. Clearly an empirical check on many such relationships is desirable to establish firmly the charge independence of the strong interactions.

The interaction between a Λ^0 particle and nucleons provides a very simple illustration. With $T = 0$ for the Λ^0 particle, exactly the same nuclear interaction should be effective in Λ^0 N and Λ^0 P systems, these being the substates of a Λ^0–nucleon $T = \frac{1}{2}$ doublet. Actually this conclusion can be derived from the less restrictive principle of charge symmetry; under the operation of charge symmetry, neutron and proton are interchanged, the Λ^0 particle remaining invariant, so that the Λ^0 N, Λ^0 P systems are charge symmetric. The restriction (2.20) also arises from charge symmetry since the neutral pion wave function reverses under the charge symmetry operation. Evidence on charge symmetry for the Λ^0 particle has recently been obtained (§ 6.2) from study of Λ^0 particles bound in nuclei.

To illustrate the nature of relationships following from the full assumption of charge independence, we will discuss the absorption of K$^-$ particles by nucleons, leading to a final $\Sigma+\pi$ state according to equation (2.25). An initial K$^-$+P system of two $T=\frac{1}{2}$ particles has $T_3=0$ and consists equally of a $T=0$ and a $T=1$ state, following expressions (2.3), nucleon 1 being replaced by a $\bar{\mathrm{K}}$ particle. Let a_0 and a_1 denote the reaction amplitudes for the states $T=0$ and $T=1$; charge independence requires these amplitudes to depend only on the total T. The final $\Sigma+\pi$ state will then be

$$\frac{1}{\sqrt{2}}[a_0\phi_{0\,0}(\Sigma\pi)+a_1\phi_{1,0}(\Sigma\pi)]$$
$$=\frac{1}{\sqrt{2}}\left[\left(\frac{a_0}{\sqrt{3}}+\frac{a_1}{\sqrt{2}}\right)\mathrm{U}_+(\Sigma)\,\mathrm{U}_-(\pi)-\frac{a_0}{\sqrt{3}}\mathrm{U}_0(\Sigma)\,\mathrm{U}_0(\pi)+\left(\frac{a_0}{\sqrt{3}}-\frac{a_1}{\sqrt{2}}\right)\mathrm{U}_-(\Sigma)\,\mathrm{U}_+(\pi)\right], \qquad \ldots\ldots(2.30)$$

using expressions (2.7) for the isotopic spin states of two $T=1$ particles. The K$^-$+P reaction therefore leads to the following charge states

$$(\Sigma^+\pi^-):(\Sigma^0\pi^0):(\Sigma^-\pi^+)=\frac{(\sqrt{2}a_0+\sqrt{3}a_1)^2}{12}:\frac{a_0^2}{6}:\frac{(\sqrt{2}a_0-\sqrt{3}a_1)^2}{12}, \qquad \ldots\ldots(2.31)$$

with the probabilities given. Since a_0 and a_1 are generally complex, the only restrictions imposed on these probabilities by the assumption of charge independence take the form of inequalities

$$\sqrt{n(\Sigma^+)}+\sqrt{n(\Sigma^-)}\geqslant 2\sqrt{n(\Sigma^0)}\geqslant|\sqrt{n(\Sigma^+)}-\sqrt{n(\Sigma^-)}|, \qquad \ldots\ldots(2.32)$$

as given by Case, Karplus and Yang (1956), the $n(\Sigma)$ denoting the numbers of Σ particles with the specified charge sign which result from the reaction considered. This inequality has now been checked against some preliminary data on the K$^-$+P reactions, which will be discussed in § 4.7.

The reaction K$^-$+N goes through $T=1$ only, since $T_3=-1$, and the final $\Sigma+\pi$ state is

$$a_1\phi_{1,-1}(\Sigma\pi)=\frac{a_1}{\sqrt{2}}[\mathrm{U}_-(\Sigma)\,\mathrm{U}_0(\pi)-\mathrm{U}_0(\Sigma)\,\mathrm{U}_-(\pi)], \qquad \ldots\ldots(2.33)$$

with the same reaction amplitude a_1 which occurs in (2.30), according to the requirements of charge independence. The final charge states therefore occur with the following probabilities

$$(\Sigma^-\pi^0):(\Sigma^0\pi^-)=\tfrac{1}{2}a_1^2:\tfrac{1}{2}a_1^2. \qquad \ldots\ldots(2.34)$$

If the K$^-$ is incident with equal probability on neutron or proton (as, for example, in interaction with ^{4}He or ^{12}C) then the number of Σ particles of given charge may be obtained by averaging the probabilities given for that charge in (2.31) and (2.34). It will then be seen that, irrespective of the values of a_0 or a_1,

$$n(\Sigma^+)+n(\Sigma^-)=2n(\Sigma^0). \qquad \ldots\ldots(2.35)$$

Although this proves to be a useful relation, the difficulty of identifying Σ^0 particles renders its verification a difficult proposition.

Feldman (1956) has given an exhaustive discussion of the various equalities and inequalities resulting from charge independence in elementary processes. Generally the relationships are complicated by the possibility of many isotopic spin channels, and the difficulty of identifying neutral particles makes many relevant reactions unobservable. The simplest predictions are those given by Lee (1955) for reactions following the absorption of K$^-$ particles by ^{4}He nuclei, since the initial system has a definite isotopic spin $T = \frac{1}{2}$. Lee points out that the isotopic spin assignments require branching ratios 2 : 1 between the following reactions (*a*) : (*b*) :

$$K^- + {}^4He \rightarrow \begin{cases} \Sigma^- + {}^3He, \\ \Sigma^0 + {}^3H, \end{cases} \text{ or } \begin{cases} \Lambda^0 + \pi^- + {}^3He, & \ldots\ldots(2.36a) \\ \Lambda^0 + \pi^0 + {}^3H. & \ldots\ldots(2.36b) \end{cases}$$

At the present time, however, no experimental evidence giving a clear-cut check on full charge independence in strange-particle interactions has been reported. On the other hand, we have no reason for disbelieving that this principle holds for strong interactions, at least to a good approximation ; no evidence of any violation of charge independence has been found for fast processes.

§ 3. LABORATORY PRODUCTION OF STRANGE PARTICLES

Since about 1953, laboratory experiments on strange-particle production using incident particle beams of well-defined energy have become possible, with the use of high-energy proton synchrotrons, the Cosmotron (maximum energy 3 GeV) of the Brookhaven National Laboratory and the Bevatron (maximum energy 6·2 GeV) of the Radiation Laboratory of the University of California, Berkeley. These machines have recently been described in these *Reports* by Blewett (1956). Their internal beam of accelerated protons may be directed to collide with internally placed target material, with a resultant production of pions and strange particles. Well-defined beams of secondary particles have been obtained from these by means of suitable magnetic focusing and analysing systems. For example, the high-energy pions emitted in the forward directions are bent by the field of the accelerator magnet ; since the negative pions are bent outwards from the machine, these have provided convenient π^- beams, after further collimation and magnetic analysis. Experiments on pion interactions in hydrogen with these beams will be described below (§ 3.1). Beams containing appreciable numbers of K mesons have been obtained by magnetic selection of secondary particles of lower momentum, as described below, and have been used in studies of K particle lifetimes, decay modes and interactions. Recently, it has been found possible to extract about one-half of the internal beam from the Cosmotron to provide a powerful focused external proton beam, which will have strong advantages for many experiments on the elementary processes. Only a small number of the possible production situations have been appreciably investigated up to the present ; the available data on these production experiments will be described briefly in the following sections.

3.1. *Production Processes in Pion–Hydrogen Collisions*

Strange-particle production in pion–proton collisions was first studied with 1·37 GeV negative pions by Fowler, Shutt, Thorndike and Whittemore (1953, 1954, 1955), using hydrogen-filled diffusion chambers. In this work nine interactions in the gas were observed to give rise to a Λ^0, θ^0 or Σ^- decay in the sensitive region of the chamber. In one interaction, both Λ^0 and θ^0 decays were found; from the momenta and angles observed for their decay products it was established that this event was due to the reaction

$$\pi^- + \mathrm{P} \rightarrow \Lambda^0 + \theta^0. \qquad \ldots\ldots(3.1)$$

This event is shown in figure 2. It provided the first clear evidence for associated production of strange particles. In the other events where only Λ^0 or θ^0 decay was observed, the mass of the unobserved neutral particle could be deduced from the observed momenta and angles, assuming a two-body production process, and each event was found compatible with this reaction (3.1). However, for several of the events, this interpretation required a rather low value to be assumed for the incident pion momentum. Fowler, Shutt, Thorndike and Whittemore (1955) pointed out that these cases could be better interpreted as Σ^0 production events,

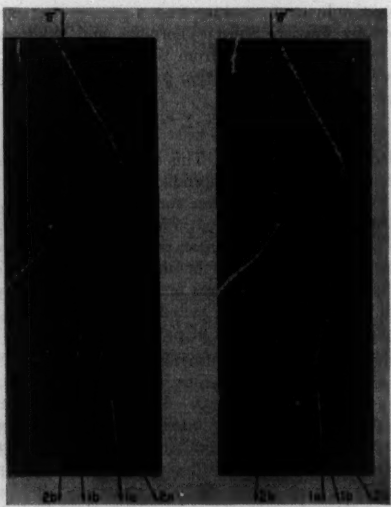


Figure 2. Diffusion cloud chamber photograph (after Fowler, Shutt, Thorndike and Whittemore 1954) of the associated production of Λ^0 and θ^0 particle in the collision of a 1·5 GeV π^- meson with a proton. Tracks 1a and 2a are due to a proton and a π^- meson, the decay products of the Λ^0 particle, and tracks 1b and 2b are due to a π^+ and π^- meson resulting from θ^0 decay.

$$\pi^- + \mathrm{P} \rightarrow \Sigma^0 + \theta^0, \qquad \ldots\ldots(3.2)$$

followed by a rapid Σ^0 decay, following Gell-Mann's suggestion (2.16). Alternatively, the incident energy is sufficiently high for some of the latter events to be the result of production leading to $\Lambda^0 + \theta^0 + \pi^0$.

Walker (1955) has somewhat clarified this situation by an experiment at 1·1 GeV. The lower energy brings two advantages. Since the incident energy is now much closer to the threshold energy (0·77 GeV) for reaction (3.1), the production of an additional neutral pion in this reaction is much less probable. Also the final particles tend to have correspondingly lower velocities, giving a greater probability for decay in the sensitive region. Two complete examples showing Λ^0 and θ^0 decay were found. For one of these, a fit to the reaction (3.1) appeared to be excluded; from the θ^0 decay measurements, a mass corresponding to a Q value of 117 ± 30 MeV in $(\pi^- + p)$ decay was obtained for the neutral baryon so that the interpretation (3.2) seemed most natural for this event.

Decay of a negative hyperon resulting from a pion–proton interaction in the gas was also observed by Fowler, Shutt, Thorndike and Whittemore (1954). Coplanarity of the two outgoing tracks from this interaction and the incident pion direction suggested a two-body production process and the measurements were in accord with the process

$$\pi^- + P \rightarrow \Sigma^- + K^+, \qquad \ldots\ldots(3.3)$$

followed by Σ^- decay in the chamber. This event provided the first clear evidence for the Σ^- particle. Fowler, Shutt, Thorndike and Whittemore (1955) and Walker and Shephard (1956) have reported three further events of charged hyperon production in a diffusion chamber without magnetic field; since the Gell-Mann scheme does not permit Σ^+ production in this reaction, these events were interpreted by reaction (3.3), although, in the absence of a magnetic field, no direct evidence on the hyperon charge sign was obtained experimentally.

[*Note added in proof*: Total cross sections for strange particle production (in mbn) in $\pi^- + P$ collisions for a number of pion energies from 0·87 to 1·1 GeV were reported at the Seventh Rochester Conference (Glaser 1957) by various groups, as follows:

π^- energy (GeV)	0·87	0·95	1·10
$\Lambda^0 + \theta^0$	$0{\cdot}6 \pm 0{\cdot}2$	$0{\cdot}45 \pm 0{\cdot}11$	$0{\cdot}30 \pm 0{\cdot}15$
$\Sigma^0 + \theta^0$	—	$0{\cdot}40 \pm 0{\cdot}10$	$0{\cdot}30 \pm 0{\cdot}15$
$\Sigma^- + K^+$	—	$0{\cdot}14 \pm 0{\cdot}07$	$0{\cdot}19 \pm 0{\cdot}10$
$\sigma_{\text{tot}}(\pi^-, P)$	48 ± 3	47 ± 3	45 ± 3

An estimate of 0·9 mbn had been obtained previously for the total strange particle production cross section for 1·37 GeV pions by Fowler, Shutt, Thorndike and Whittemore (1955)—at this energy the total (π^-, P) cross section is about 34 mbn.]

Now the angular correlations which may occur in the reaction (3.2) and in the subsequent decays are of great interest, owing to their dependence on the mechanism of the reaction and on the strange-particle spin values. However, there are very many correlations which may be examined, and the various angles which may be recorded are defined as indicated in figure 3. The distribution in the production angle θ_K will have a direct dependence on the production mechanism. The correlations of the direction (θ, ϕ) of a final decay pion (seen in the rest system of its parent particle and measured from the line OK in the initial (π^-, P) centre-of-mass system and from the production plane POK), with the direction of motion of the parent particle, with the production plane and with the direction of decay of the accompanying strange particle will each depend strongly on the spin–parity values for the strange particles involved, as well as on the production mechanism.

The most significant remark on these correlations has been that made by Adair (1955), who noted that since the initial system has magnetic quantum number $\pm 1/2$ relative to the incident direction, this must also apply for the final Λ or Σ hyperon (assuming spin zero for the K particle) if it is emitted with $\theta_K = 0$ or π, irrespective of its spin value. This means that, for hyperons emitted forwards or backwards, the angular distribution in their decay is uniquely determined by their spin value, being isotropic for a hyperon spin 1/2, $(1+3\cos^2\theta_1)$ for hyperon spin 3/2, and so on. Of course, this result will hold to a good approximation also for emission with finite angles θ_K provided that the contributions corresponding to hyperons carrying off non-zero orbital angular momentum component m_l are sufficiently small (this requires angles $\theta_K \lesssim 1/l$ or $\pi - \theta_K \lesssim 1/l$, where l denotes the highest outgoing angular momentum giving significant contributions in the process).

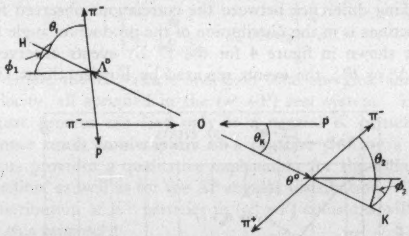


Figure 3. The angles of interest in the reaction $\pi^- + P \rightarrow \Lambda^0 + \theta^0$ and the subsequent Λ^0 and θ^0 decays. The direction of motion (and the angle θ_K) shown for the Λ^0 and θ^0 is that specified in the $\pi^- + P$ centre-of-mass system, the decay lines (π^+, π^-) and (π^-, P) for θ^0 and Λ^0 decay are those observed in the θ^0 and Λ^0 rest systems respectively.

The most extensive information on these correlations is that reported recently by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956), who exposed a propane bubble chamber to a 1·43 GeV π^- beam at Brookhaven. As no magnetic field was available for this work, all events have been analysed on the assumption that they represent reactions (3.1), (3.2) and (3.3), as required by the Gell-Mann rules. On this basis, the interactions occurring in the hydrogen may be separated from the carbon interactions in the chamber by criteria depending on the kinematics expected for these reactions. Thus, to identify reaction (3.3), it is required that the three tracks be coplanar, that the angles satisfy the kinematical relation expected for this reaction, that the observed bubble densities and multiple scattering should match the energies expected for the observed angles and that either K^+ or Σ^- should decay in the chamber. Similarly, for reaction (3.1), the $\Lambda^0(\theta^0)$ decay plane is required to contain the point of interaction and the angles of the decay products must be in accord with the velocity expected for a $\Lambda^0(\theta^0)$ produced in the direction observed.

In a Σ^0 reaction (3.2), the Σ^0 particle will decay extremely close to its origin, so that the event will have a general appearance very similar to the Λ^0 reaction (3.1).

However, the energies and angles of the resulting Λ^0 and θ^0 will then usually be somewhat different from those for the Λ^0 reaction; the Σ^0 interpretation (3.2) does seem necessary in two of the cases where only the θ^0 is seen to decay. When both Λ^0 and θ^0 decay are seen the observations provide a number of checks by which reactions (3.1) and (3.2) may be separated. One such event (out of four observed) appears to fit the Σ^0 reaction. But since ambiguities arising from the carbon contamination are possible, it has not been claimed that these observations, taken alone, provide any strong argument for the existence of the Σ^0 particle.†

The most striking difference between the correlations observed for the (Σ^-K^+) and the ($\Lambda^0\theta^0$) reactions is in the distribution of the production angle θ_K. Distributions in cos θ_K are shown in figure 4 for the 17 Σ^- events observed and for the 38 events giving Λ^0 or θ^0; the events reported by Fowler, Shutt, Thorndike and Whittemore (1955) are also shown (dotted lines). The result is that, in the laboratory system, Σ^- particles tend to be produced with high velocity, whereas the Λ^0 particles, being produced backwards in the centre-of-mass systems, are predominantly of low velocity. Alternatively this may be expressed by saying that the K^0 mesons produced are projected strongly in the forward direction, whereas this is not the case for the K^+ mesons produced. The angular distribution $(1+\cos\theta_K)^2$ shown in figure 4 gives the greatest forward peaking possible if only s and p waves are effective in the final system. This comparison suggests that still higher orbital angular

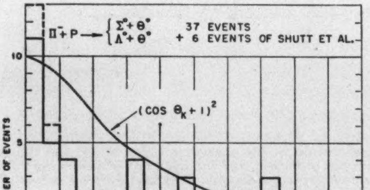


Figure 4. The distribution observed (after Budde, Chretien, Leitner, Samios, Schwartz and Steinberger 1956) for the production angle θ_K in the reactions $\pi^- + P \rightarrow \Lambda^0 + \theta^0$ and $\pi^- + P \rightarrow \Sigma^- + K^+$. The events of Fowler *et al.* (not Shutt *et al.*, as in the figure) are shown by the dotted lines (the sixth event in the upper diagram should be inserted between 0·8 and 0·7).

---

† *Note added in proof*: Plano, Samios, Schwartz and Steinberger (1957) have recently reported three ($\pi^- + P$) events in which both decay products of the Σ^0 particle have been observed, the γ-ray producing an electron–positron pair within the chamber. This demonstrates very clearly the existence of the Σ^0 particle since these events give a unique value $Q(\Sigma^0 \rightarrow \Lambda^0 + \gamma) = 73{\cdot}0 \pm 3{\cdot}5$ Mev from measurements of the Λ^0 and γ-ray momenta. This corresponds to a total energy of ($\pi^- + P + 110 \pm 3{\cdot}5$ Mev) so that the Σ^0 mass is rather close to the Σ^+ mass.

momenta must contribute quite strongly in the final state, even though the wavelength of the outgoing particles is of the order of 10^{-13} cm at this energy.

One possible explanation for these striking features in K particle production from $(\pi^- + \mathrm{P})$ collisions has been pointed out by Barshay (1956) and by Schwinger (1956), who propose that the interaction for absorption of a pion by a K particle should be a strong interaction (cf. § 7.4). An angular distribution strongly peaked in the forward direction may then result from capture of the incident π^- meson by the outgoing K particle, for the angular distribution due to this effect will be dominated by a factor

$$\left\{1 - \frac{m_\pi^2}{2E_\pi E_\mathrm{K}} - \beta_\mathrm{K} \cos\theta_\mathrm{K}\right\}^{-2},$$

where E_K, E_π denote the K particle and pion total energies, and β_K the outgoing K particle velocity, all specified in the $(\pi^- + \mathrm{P})$ rest system. Further, it is clear that this capture process can lead only to a neutral K particle since a positive K particle cannot result from capture of a negative pion by a K particle. This mechanism thus provides a qualitative explanation for the difference between K^0 and K^+ production, as well as for the K^0 angular distribution. The observation of the angular distribution of K^+ particles in $(\pi^+ + \mathrm{P})$ collisions will provide an interesting test for this proposal.[†]

Correlations involving the decay processes were also examined but none of these were significantly different from isotropy. At the present time therefore, the correlations in these processes provide no evidence that the hyperon spin is greater than $\frac{1}{2}$, nor that the θ^0 spin is non-zero, although the absence of correlations does not exclude these possibilities. The application of Adair's procedure (see above) to the limited data available[‡] gives no indication of higher spin values for the Λ^0 particle.

Table 1. Threshold Kinetic Energies for Various K Particle Production Processes due to Nucleons or Pions Incident on Hydrogen or Nuclear Targets

Nucleon–nucleon collisions leading to final products	Hydrogen target (GEV)	Nuclear target (Fermi momenta up to 220 MEV/c) (GEV)
$\Lambda + \mathrm{K} + \mathcal{N}$	1·57	1·10
$\Sigma + \mathrm{K} + \mathcal{N}$	1·80	1·27
$2\mathcal{N} + \mathrm{K} + \overline{\mathrm{K}}$	2·50	1·81
$\Xi + 2\mathrm{K} + \mathcal{N}$	3·74	2·79
$\Lambda + \Lambda$	0·77	0·49
$\Sigma + \Sigma$	1·16	0·78
Pion–nucleon collision leading to final products		
$\Lambda + \mathrm{K}$	0·76	0·58
$\Sigma + \mathrm{K}$	0·90	0·69
$\mathcal{N} + \mathrm{K} + \overline{\mathrm{K}}$	1·36	1·05
$\Xi + 2\mathrm{K}$	2·23	1·73

[†] *Note added in proof*: Velde, Brown, Glaser, Meyer, Perl, Cronin, De Benedetti and Cool (1957) have recently observed the reaction $\pi^+ + \mathrm{P} \rightarrow \Sigma^+ + \mathrm{K}^+$. The total cross section is $0{\cdot}13^{+0{\cdot}12}_{-0{\cdot}05}$ mbn at 1·1 GEV. The angular distribution of the Σ^+ particles has a strong forward peak (similar to Σ^- particles in the $\pi^- + \mathrm{P}$ reaction); this disagrees with expectation from the above arguments.

[‡] *Note added in proof*: The application of Adair's procedure to the quite extensive data on Λ and Σ production reported at the 1957 Rochester Conference (an increase in statistics by a factor of about ten) now points very strongly to spin 1/2 for the Λ and Σ hyperons, if spin zero is assumed for the K mesons.

3.2. *Gamma Ray Observations on Strange Particle Production*

Strange particles produced in a target of moderate dimensions by proton bombardment will generally decay at appreciable distances outside the target. Those particles whose decay products include a neutral pion will therefore provide a source of γ rays in the neighbourhood of the target. Observations of such γ rays were first made at Brookhaven by Ridgway, Berley and Collins (1956). The apparatus consisted of a well-collimated γ-ray telescope, sensitive only to γ rays exceeding 10 MeV. An external copper target was placed upstream from the region explored by the telescope and the γ-ray intensity was measured as a function of the distance x upstream to the target.

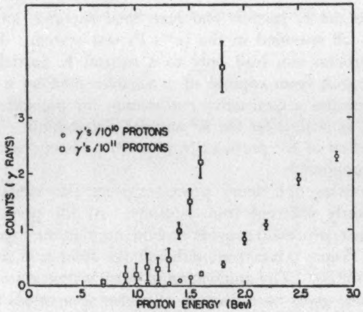


Figure 5. The excitation function for γ rays produced more than 1·3 cm downstream from a copper target 0·94 cm thick bombarded by protons (after Ridgway, Berley and Collins 1956).

For fixed proton energy (in the region of 2 GeV), the photon intensity was found to decrease exponentially with $x(\sim e^{-x/d})$, the characteristic distance d being 4·16($\pm$0·1) cm at 2·9 GeV. For a fixed distance x in this exponential region, the dependence of the photon intensity on proton energy was measured, as shown in figures 5 and 6. The photon intensity drops rapidly with proton energy, becoming indistinguishable from background below 1·1 GeV. After taking into account the momentum distribution of the nucleons within the target nuclei, due to their Fermi motion, the threshold (laboratory energy) for the elementary nucleon–nucleon collision is found to be about 1·5 GeV. This is quite close to the threshold of 1·57 GeV expected for associated production of the type

$$P+N\rightarrow\Lambda^0+K^0+P, \qquad \ldots\ldots(3.4)$$

which has the lowest threshold of all the strange-particle production processes allowed by charge independence for nucleon–nucleon collisions (cf. table 1). Assuming the cross section for (3.4) to be proportional to the phase space available to the final particles, and taking into account the Fermi motion of the target nucleons, a very good fit is obtained to the observed excitation function, as shown in figure 6.

The interpretation of the characteristic distance d in terms of a lifetime does not depend very sensitively on the centre-of-mass distribution of the final particles in (3.4). The lifetime obtained was $1{\cdot}4 \times 10^{-10}$ sec, with variations of order 10% with various such assumptions. Since the only known lifetime compatible with this is the θ^0 lifetime, it has been concluded that the photons observed result from an alternative decay mode,

$$\theta^0 \rightarrow \pi^0 + \pi^0. \qquad \ldots\ldots(3.5)$$

With the higher proton energies available at Berkeley, a more extensive experiment has been possible on the same lines; a preliminary report has been given by Moyer, Osher and Parker (1956). In this experiment the γ-ray telescope can move upstream and downstream relative to the target. Figure 7(*a*) shows a typical distribution of photon intensity at the highest proton energy. The central peak is due to the decay of π^0 mesons formed directly in the target, but the presence of exponential decay curves both upstream and downstream from the target is clearly shown. The upstream component is of particular interest since only K particles can travel upstream from the target. In the downstream component there appears two exponential decay curves, as shown in figure 7(*b*), the intensity of the longer-lived component falling rapidly as the proton energy falls below 3 GeV. The preliminary analysis of these decay curves again gives $1{\cdot}5 \times 10^{-10}$ sec for the K meson lifetime involved; the longer-lived downstream decay corresponds to a lifetime consistent with the Λ^0 lifetime and is believed due to the alternative neutral decay,

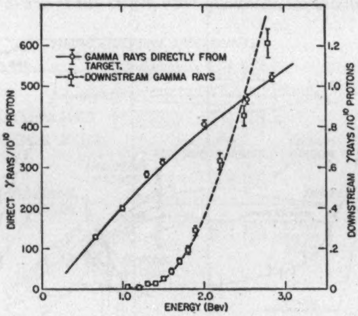


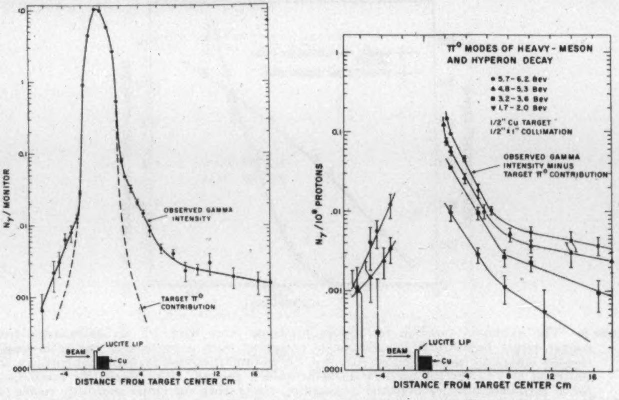
Figure 6. The excitation function for γ rays produced more than 3·7 cm downstream from a copper target bombarded by protons is compared with a theoretical energy dependence (broken curve) for production of $(\mathrm{K}+\Lambda+\mathcal{N})$, assuming s-states for the final particles and a threshold of 1·53 GeV for the free nucleon–nucleon process. Also shown for comparison is the π^0 excitation function obtained by counting γ rays when the target is directly visible (after Ridgway, Berley and Collins 1956).

$$\Lambda^0 \rightarrow \mathrm{N} + \pi^0. \qquad \ldots\ldots(3.6)$$

These experiments establish a number of important results. Among these we will mention especially the following:

(i) The existence of the Λ^0 decay (3.6) and the absence of photons below 1·1 GeV proton energy imply that nucleon–nucleon processes such as (2.9), which lead to two hyperons alone, can have cross sections at most several orders of magnitude smaller than processes of the type (3.4), despite their considerably lower thresholds.

(ii) The decay $\theta^0 \rightarrow \pi^0 + \pi^0$ to two identical bosons implies that the θ^0 particle must have even parity. This may be seen from the fact that a parity operation in the final state reverses the relative position vector between the two pions, a change which is equivalent to the interchange of these neutral pions. Since the neutral pions are identical bosons, the final state must be unaffected by their interchange, and therefore also by the parity operation, i.e. the parity of the final state must be even. Further, the parity of the final state arises only from the relative orbital motion of the two pions, so that it follows that this orbital angular momentum must be even; since pions are spinless, the θ^0 spin value is therefore necessarily even. To sum up, the existence of the decay mode (3.5) implies that the only spin–parity values possible for the θ^0 particle are limited to 0^+, 2^+, etc.



(*a*) (*b*)

Figure 7. (*a*) The γ-ray intensity upstream and downstream from a $\frac{1}{2}$-in. copper target bombarded by 5·7–6·2 GeV protons, showing the θ^0 decay curve upstream from the target and well-resolved Λ^0 and θ^0 decay curves downstream. The central peak is due to γ rays from π^0 mesons directly produced in the target. (*b*) The decay curves observed for γ rays produced upstream and downstream from a copper target are shown as functions of the bombarding proton energy.

3.3. *The Production of Charged K Particle Beams*

The beams of particles which have been used for the majority of K particle investigations are obtained from proton bombardment of internal targets (usually Cu or Ta). The secondary beam leaving the accelerator through a thin window passes through a strong-focusing magnetic lens, thence through a collimator defined by lead walls and finally through a bending magnet by means of which the final momentum of the beam is defined. Figure 8 shows the arrangement first developed at the Bevatron by Kerth and Stork (Birge, Perkins, Peterson, Stork and Whitehead 1956). In this arrangement the secondary beam is taken out at 90° to the proton beam direction: most of the K particle experiments have used momenta in the neighbourhood of 400 MeV/c.

The composition of a typical positive beam (375 MeV/c) has been investigated by Smith, Heckman and Barkas (1956) for 4·6 GeV protons (Ta target). This contained 87·3% protons, 3·8% pions, 0·8% positrons and 0·04% K^+ mesons, the remainder of the beam consisting of ^2H, ^3H, ^3He and ^4He nuclei. When this beam passes into matter, the protons and heavier particles are quickly brought to rest, the positrons are degraded and a beam of K^+ particles and π^+ mesons is left. The K^+/π^+ ratio is then about 1–2% at the end of the arrangement (corresponding to a K^+/π^+ ratio of order 3–5% at the target, after allowance for attenuation and decay).

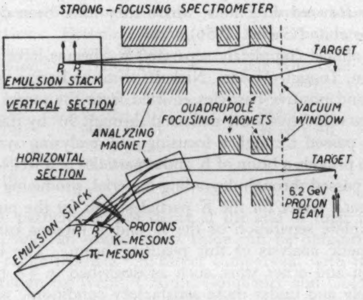


Figure 8. Diagram of the strong-focusing spectrometer arrangement used for emulsion exposures at the Bevatron (after Birge, Perkins, Peterson, Stork and Whitehead 1956).

Since K and π particles in the beam have rather different velocities, the incident particle may be identified in counter experiments by arranging for the beam to pass through a Cerenkov detector before reaching the experimental set-up. When the beam is passed into an emulsion stack, the K particles, all having a definite momentum, are brought to rest in a definite region of the emulsion, where their decay properties may be studied; the pions, with the same momentum but higher velocities, have a much longer range than the K particles and pass well beyond this region, so that they cause little confusion. Information on the variation of K^+ intensity in this beam (at 250 MeV/c) with Bevatron proton energy has been obtained by Goldhaber (1956) in emulsion studies; this intensity rises by a factor close to 3 between 4·2 and 6·2 GeV.

A similar arrangement at Brookhaven has been described by Harris, Orear and Taylor (1956). Partly owing to the longer time of flight for the particles, due to the geometrical arrangement, and also because of the lower proton energy available, the K^+/π^+ ratio reported in this beam (375 MeV/c) is about 1/600, which corresponds to a ratio of about 1% at the target. This beam has been used for emulsion studies, and especially for measurements on K^+ particle lifetimes (cf. § 4.2).

The corresponding negative beam at Berkeley consists mainly of π^- mesons. The intensity of K^- particles is very low, a K^-/π^- value of about $0{\cdot}8 \times 10^{-4}$ being given by Fry, Schneps, Snow and Swami (1955 b) for the 250 MeV/c beam at full energy. The smallness of the K^-/K^+ ratio observed (of order several per cent) is generally attributed to the high threshold (2·50 GeV) for K^- production through pair creation processes of the type (2.11), (the only K^- processes the Gell-Mann–Nishijima scheme allows at these energies) in comparison with the threshold of 1·57 GeV for production of a K^+ particle in association with a Λ^0 hyperon. This effect is naturally accentuated by the large angle (90°) at which these beams are taken out from the Bevatron.

At Brookhaven K^- particles are extremely rare in the secondary beam arrangement. Although the Brookhaven proton energy is sufficiently high for K particle pair creation, the final particles are left with rather little kinetic energy in the centre-of-mass system. The K^- particles produced can be emitted with appreciable probability only in very forward directions, where they have been detected in emulsion studies by Hornbostel and Salant (1956).

Very recently, almost completely separated $K^\pm$ beams have become available at Berkeley (Heckman, Inman, Mason, Nickolls, Smith, Barkas, Dudziak, and Giles 1956). Negative and positive particles emitted from the target in the forward and backward directions respectively are deflected through 90° by the Bevatron magnetic field and are then passed through a focusing and analysing system similar to that described above, to provide a beam of K and π particles of a well-defined momentum. This beam is next passed through degrading material, producing a mean momentum change which is much larger for the K particles than for the pions. Consequently an essentially complete separation of the K and π mesons can then be achieved by a further magnetic analysis of this resulting beam. The use of these beams will allow emulsion and other work such as described in § 4 below to be carried out more extensively and under more satisfactory conditions, with greatly reduced labour.

The distinction between K^+ and K^- particles is one of the most striking features of the Gell-Mann–Nishijima proposals. Their particle–antiparticle relationship leads to the prediction that, at present energies, K^- particles can be produced only through K pair creation processes, whilst K^+ particles can also be produced by the (Λ^0+K^+) processes effective at lower energies. The contrast between K^+ and K^- production discussed above is in good accord with this prediction and consequently gives strong support for the strangeness assignments of this scheme.

§ 4. THE CHARGED K PARTICLES AND THEIR INTERACTIONS

4.1. *Mass Measurements for Charged K Particles*

Direct measurements of K particle masses have been made using the momentum–range method by which our most accurate knowledge of pion masses (Smith, Birnbaum and Barkas 1953) had previously been obtained. When the momentum-analysed secondary beam described in § 3.3 is passed into nuclear emulsion, the various K particles come to rest in a narrow band across the emulsion since they all have about the same energy, their masses being very similar. Each K particle track may then be classified according to the decay mode observed. From the K particle energy at entry to the emulsion, obtained from its measured range in the emulsion, and the K particle momentum corresponding to its point of entry, obtained from the geometry and the known magnetic fields applied, a mass value may be deduced for each K particle. The distribution in mass values obtained for a particular K particle mode will be due to the straggling in the K particle range and to the spread in initial momenta due to the target size and to multiple scattering in the air between the target and collimator. The values obtained by Heckman, Smith and Barkas (1956 a) and Birge, Perkins, Peterson, Stork and Whitehead (1956) for the various K modes are collected in table 2; these results provide no evidence for any mass differences between the K particles and are all consistent with a mass of 966(±1·0)m_e.

Table 2. Mass Measurements for K Particles by Various Berkeley Groups. The Values of Heckman *et al.* are normalized to a τ Mass of 966m_e.

	Primary Mass (Momentum–Range)		Secondary Mass (Energy Release)	
	Birge, Perkins, Peterson, Stork and Whitehead (1956)	Heckman, Smith and Barkas (1956 a)	Birge, Perkins, Peterson, Stork and Whitehead (1956)	Barkas, Heckman and Smith (1956)
$K_{\mu 2}$	967·5 ± 2·2	962·5 ± 3·2	964·8 ± 2·2	963·4 ± 2·1
$K_{\pi 2}$	966·5 ± 2·0	972·2 ± 4·7	964·2 ± 2·0	962·6 ± 1·6
τ	966·3 ± 1·9	966 ± 3·1	965·8 ± 1·0	966·3 ± 0·7
$K_{\pi 3}$	—	951·5 ± 10·9	—	—
$K_{\mu 3}$	967·3 ± 5·5	951·2 ± 10·6	—	—
K_{e3}	963 ± 10	—	—	—

Since negative K particles do not decay at the end of their range but undergo nuclear interactions, only an overall mass value can be obtained for K^- particles by the momentum–range method. A direct comparison with the mean K^+ mass has been carried through by Webb, Chupp, Goldhaber and Goldhaber (1956) with the result

$$m(K^-)/m(K^+) = 0{\cdot}998 \pm 0{\cdot}013.$$

Independent mass values may also be obtained from measurement of the energy released in decay, at least for the two-body decay modes $K_{\mu 2}$ and $K_{\pi 2}$, and for the τ decay where all the decay products are charged. Following recent measurements by Barkas, Barrett, Cuer, Heckman, Smith and Ticho (1956) which establish points on the range–energy curve for emulsion up to pion energies of order of 150 MEV, mass values have been obtained for these modes particularly by Barkas, Heckman and Smith (1956) and Birge, Perkins, Peterson, Stork and Whitehead (1956), as listed in table 2. For the τ decay, the Q value quoted by Barkas, Heckman and Smith is $Q = 75{\cdot}08(\pm 0{\cdot}20)$ MEV; the most accurately determined K mass is therefore the τ mass, $m(\tau^+) = 966{\cdot}3(\pm 0{\cdot}7)\,m_e$, for which the error quoted arises mostly from the uncertainty in the pion masses.

4.2. *The Two-body K Particle Decays*

4.2.1. *The* $K_{\mu 2}$ *particle.*

From the momentum–range determination of the $K_{\mu 2}$ mass and the unique energy observed for the secondary meson, it may be concluded that the mass of the accompanying neutral particle cannot exceed about 70 m_e, although it can be zero. The decay therefore involves a photon or a neutrino. From the absence of photon showers in the cloud-chamber events (Gregory, Lagarrigue, Leprince-Ringuet, Muller and Peyrou 1954) by which the $K_{\mu 2}$ particle was first established, it seems clear that the decay mode involves a neutrino,

$$K^+_{\mu 2} \rightarrow \mu^+ + \nu. \qquad \ldots\ldots(4.1)$$

This conclusion is also supported by the absence of $K^+ \rightarrow \mu^+ + e^+ + e^-$ events in emulsions, which would be expected with a relative frequency of approximately 1% due to internal conversion if a photon accompanied the muon.

Electronic measurement of the $K^+_{\mu 2}$ lifetime has recently been reported by Fitch and Motley (1956) and by Alvarez, Crawford, Good and Stevenson (1956), using refinements of the method of Mezetti and Keuffel (1954), who first obtained a $K^+_{\mu 2}$ lifetime in cosmic-ray work. The arrangement of Fitch and Motley is shown in figure 9. The secondary beam passes first through Cerenkov counters C_1 and C_2; C_1 is sensitive to a limited velocity range $(0{\cdot}62 < \beta < 0{\cdot}78)$ whilst C_2 is sensitive only to velocities rather greater than the K velocity in the secondary beam. A count from C_1 without accompanying count from C_2 can arise therefore only from the passage of a K particle. The absorber A slows down the K particles so that they stop in the Cerenkov detector C_6 after passing through scintillators S_1 and S_2. C_6 is sensitive only to the relativistic secondary of the K meson decay. Since absorbers D and E are made sufficient to stop the pion secondary from $K_{\pi 2}$ decay, a count C_6 accompanied by a count C_3 or C_4 can only result from a $K_{\mu 2}$ secondary. A coincidence $C_1\overline{C}_2S_1S_2\overline{S}_3$ therefore denotes the arrival of a K particle and the time interval between this and a coincidence C_6 (C_3 or C_4) is then recorded. About 300 $K_{\mu 2}$ decays have been detected in this way and the distribution of their decay times is well fitted by an exponential curve corresponding to a mean lifetime,

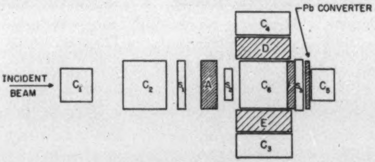


Figure 9. Counter and absorber arrangement for the measurement of K+ particle lifetimes (after Fitch and Motley 1956).

$$T(K^+_{\mu 2}) = 1{\cdot}17(\pm 0{\cdot}08) \times 10^{-8} \text{ sec.} \qquad \ldots\ldots(4.2)$$

The value obtained in a similar way by Alvarez, Crawford, Good and Stevenson is in accord with this more accurate value. (*Note added in proof*: A more accurate value of $1{\cdot}24 \pm 0{\cdot}018 \times 10^{-8}$ sec has recently been reported by these latter authors at the 1957 Rochester Conference.)

4.2.2. *The* $K_{\pi 2}$ *particle.*

The unique range of the pion secondary in $K_{\pi 2}$ decay corresponds to a pion energy close to 109 MeV (Barkas, Heckman and Smith 1956). From this energy and the $K_{\pi 2}$ mass obtained from the momentum–range measurements, it follows that the accompanying neutral particle has π^0 mass, to within $10m_e$, so that

$$K^+_{\pi 2} \rightarrow \pi^+ + \pi^0. \qquad \ldots\ldots(4.3)$$

For the photons from the subsequent $\pi^0 \rightarrow 2\gamma$ decay, there should be a definite relationship between energy and angle of emission relative to the π^+ direction, depending on the known velocity of the π^0 meson. Photons with energy and angle satisfying this relation have been observed for $K_{\pi 2}$ events in multiplate cloud chambers (Courant 1955). This relation should also hold for an electron pair from the subsequent decay $\pi^0 \rightarrow \gamma + e^+ + e^-$ and this has been verified in a number of events (e.g. Anderson, Lawlor and Nevin 1955 b, Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard 1956 a). Finally this mode (4.3) is also able to account for the K^+ decay reported by Hodson, Ballam, Arnold, Harris, Rau, Reynolds and Treiman (1954), which led to five charged particles, consisting of two electron pairs and a pion or muon. The momenta observed for these electron pairs strongly suggested that they resulted from a rare mode of π^0 decay $\pi^0 \rightarrow e^+ + e^- + e^+ + e^-$; the corresponding π^0 velocity and the momentum observed for the charged secondary then fitted well with the interpretation (4.3).

If the $K_{\pi 2}$ particle has spin j, then, since the pions are spinless, it is clear that the relative motion of the two pions must have its orbital angular momentum equal to j. The parity of this orbital motion is necessarily $(-1)^j$. Since the two pions each have parity (-1), it follows that this decay is possible only if the K particle has intrinsic parity $(-1)^j$. Further, it is natural to associate the $K^+_{\pi 2}$ with the θ^0 particle, for which the neutral decay mode (3.5) requires an even parity. With this restriction only spin–parity values $0^+, 2^+, 4^+, \ldots$ are possible for the $K_{\pi 2}$ particle.

The $K^+_{\pi 2}$ lifetime has also been measured electronically. In the arrangement of Fitch and Motley (see figure 9), absorber A is modified so that the K particles stop in the absorber F. The $K^+_{\pi 2}$ particle is then identified by a coincidence between the Cerenkov detector C_6 and the detector C_5 behind the lead converter. C_6 does not respond to secondaries from $K_{\pi 3}$ or τ decay, its velocity threshold being $\beta = 0{\cdot}7$; on the other hand some K_{e3} events will be counted together with the much more frequent $K_{\pi 2}$ events. The events recorded lay on an exponential curve of mean lifetime

$$T(K^+_{\pi 2}) = 1{\cdot}21(\pm 0{\cdot}11) \times 10^{-8} \text{ sec.} \qquad \ldots\ldots(4.4)$$

This is in agreement with the lifetime obtained in a similar experiment by Alvarez, Crawford, Good and Stevenson (1956) at Berkeley. (*Note added in proof*: These latter authors have reported a value $1{\cdot}21 \pm 0{\cdot}03 \times 10^{-8}$ sec at the 1957 Rochester Conference.)

4.3. *The τ Particle*

The characteristic appearance of τ^+ decay in emulsion, giving three coplanar charged pions each with a ' grey ' or ' black ' track, allows τ^+ events to be identified and measured quite rapidly. Frequently all three pions may be followed to the end of their range, establishing the configuration of the decay event in complete detail. However, even if only two pions may be followed to rest, the configuration may still be deduced, since the third pion energy may be deduced from these observed energies and the angles between the pion tracks by requiring conservation of momentum, the interpretation being checked by the requirement that the resulting Q value be in accord with the known mean value of 75·1 MEV. To the present time nearly 900 τ^+ decay configurations have been measured by various groups† in emulsions exposed to the Bevatron secondary beam.

Further, the τ^+ decay provides an especially favourable situation for analysis; the products consist of three equally massive particles, two of them identical, with all momenta known. The configuration of the decay may be conveniently specified in terms of two momenta; $\mathbf{p}$, the momentum of the negative pion in the initial rest system, and $\mathbf{q}$, the momentum of one of the two positive pions in their centre-of-mass system (the other positive pion naturally having momentum $-\mathbf{q}$ in this system). The magnitudes of these two momenta are related to the primary τ^+ mass M_τ by

$$q = (M_\tau^2 - 3m_\pi^2 - 2M_\tau \omega_{\mathrm{p}})^{\frac{1}{2}}/2, \qquad \ldots\ldots(4.5)$$

where $\omega_{\mathrm{p}} = (m_\pi^2 + p^2)^{\frac{1}{2}}$ is the total π^- energy. If no account is taken of the spatial orientation of the event, it suffices to specify the π^- energy $\omega_{\mathrm{p}} = m_\pi + t$ (t being the π^- kinetic energy) and the angle θ between the vectors $\mathbf{p}$ and $\mathbf{q}$; the latter may be calculated from the observed momenta by the expression

$$\begin{aligned} \cos\theta &= \frac{p_1^2 - p_2^2}{pq}\left\{1 - \left(\frac{p}{M_\tau - \omega_{\mathrm{p}}}\right)^2\right\}^{-\frac{1}{2}}, \\ &= \frac{\omega_1 - \omega_2}{p}\,\frac{\omega_{\mathrm{q}}}{q}, \end{aligned} \qquad \ldots\ldots(4.6)$$

p_1, ω_1 and p_2, ω_2 being the observed (laboratory) momenta and energies for the two positive pions ($p_1 > p_2$). Note that $p/(M_\tau - \omega_{\mathrm{p}})$ is the velocity of the $\pi^+ - \pi^+$ rest system relative to the initial rest system. The advantage in this choice of coordinates is that the relation (4.5) between p and q is independent of $\cos\theta$.

Next, we denote by L, l the angular momenta associated with the motion described by each of the vectors $\mathbf{p}, \mathbf{q}$ respectively. Since the positive pions are identical Bose particles, the motion must be unaffected by their interchange, which corresponds to reversal of $\mathbf{q}$; this is possible only if their relative motion has even parity, and therefore even l. The parity of the system is consequently $(-1)^L \times (-1)^3$, taking into account odd intrinsic parity for each pion. Conservation of parity in the τ decay therefore requires that

$$w = -(-1)^L, \qquad \ldots\ldots(4.7)$$

where w denotes the intrinsic parity of the τ particle. It follows that only even or odd L are permitted according as the τ parity w is even or odd. Finally, these angular momenta L, l must add vectorially to give the τ spin j,

$$\mathbf{j} = \mathbf{L} + \mathbf{l}. \qquad \ldots\ldots(4.8)$$

For zero spin, $j = 0$, it is clear that angular momentum conservation requires that L and l be equal and therefore both even, so that, from (4.7), the total parity of the three pions is necessarily negative. It follows that decay of a particle of spin and parity 0^+ to three pions is forbidden since it cannot conserve parity. If the τ^+ and $K_{\pi 2}^+$ particles are identical, then, following § 4.2.2, this particle can only have spin-parity 2^+, 4^+, etc., assuming parity conservation to hold for these decay processes.

† These τ^+ events consist of the following: 386 events by Baldo-Ceolin, Bonetti, Greening, Limentani, Merlin and Vanderhaege (1957), 191 events by Haddock (1956 and private communication), 106 events by Biswas, Ceccarelli-Fabbrichesi, Ceccarelli, Gottstein, Varshneya and Waloschek (1956 a), 72 events by Orear, Harris and Taylor (1956), 33 events by Heckman, Smith and Barkas (1956 b), 33 events (line scan) by Bhowmik, Evans, van Heerden and Prowse (1956), 23 events (unbiased) by Brene, Hansen, Hooper and Scharff (1956) and 49 events by Feld, Odian, Ritson and Wattenberg (1955).

For all non-zero spins and either parity, there exists an infinite sequence of sets (L, l) for which (4.7) and (4.8) are satisfied. However, several rigorous statements can be made generally:

(i) If the τ^+ particle has either even parity w or odd spin j, its decay cannot give rise to an s-wave π^- meson. This follows from (4.7) and (4.8) since $L = 0$ requires $w = -1$ and $j = l =$ even. Thus, in particular, *the probability density for* τ^+ *decay events (per unit phase space) must vanish for the decay configuration in which the* π^- *meson is at rest, unless the* τ *parity is odd and the* τ *spin is even.*

(ii) Two π^+ mesons from τ^+ decay cannot have relative s-wave motion (but must be in at least a relative d-wave) if its spin and parity satisfy $w = (-1)^j$. Since $l = 0$ requires $j = L$, (4.7) implies that relative s-wave motion is possible only for $w = -(-1)^j$. From (4.5), a low relative energy between the π^+ particles implies a π^- energy correspondingly close to the maximum value possible. *With* $w = (-1)^j$, *the probability density for* τ *decay events must vanish for the decay configuration in which the* π^- *meson has its maximum possible energy.*

(iii) A configuration in which the three pions have parallel momenta (so that $\cos\theta = 1$) cannot result from τ^+ decay if $w = (-1)^j$. In this case only one direction is available to specify the configuration, whose orbital motion, for total angular momentum j, must then be described by the spherical harmonic Y_j^m of this direction; consequently the total parity of this configuration is necessarily $(-1)^3 \times (-1)^j$. *With* $w = (-1)^j$, *the probability density for* τ *decay events must vanish for configurations with* $\cos\theta = 1$.

These three remarks provide the essential basis for distinguishing between various spin–parity possibilities for the τ^+ particle, from observations on τ^+ decay configurations. These are three situations in which the probability density must vanish if $w = (-1)^j$, which is the relationship which must hold for the $K_{\pi 2}$ particle (cf. § 4.2.2). For other spin–parity values, there may be additional restrictions and these will be mentioned in the discussion below.

The probability distribution for τ decay configurations consists essentially of factors (*a*) the phase space density ρ for the configuration,

$$\rho\, d\omega_D\, d(\cos\theta) = \{pq/(M_\tau^2 + m_\pi^2 - 2M_\tau \omega_D)^{\frac{1}{2}}\}\, d\omega_D\, d(\cos\theta)/(2\pi)^6, \qquad \ldots\ldots(4.9a)$$

and (*b*) the probability density Φ, which is the square of the amplitude $S(m)$ for this decay configuration, averaged over τ spin orientations,

$$\Phi = \frac{1}{2j+1} \sum_m |S(m)|^2. \qquad \ldots\ldots(4.9b)$$

This second factor Φ is the quantity of interest, whose properties are closely related with the τ spin–parity. Empirically this may be obtained by dividing the observed density of events by the phase space density ρ. In table 3 this has been done for the 495 events on which sufficiently detailed data are at present available.

The result shows no strong systematic dependence on $t = \omega_D - m_\pi$ and $\cos\theta$ but is approximately uniform within statistical fluctuations; there is no indication that the probability distribution goes to zero as the decay configuration approaches any of the limiting situations mentioned above. This is emphasized in figures 10 and 11 which show the distributions directly bearing on these situations, the distributions over π^- energies and over $\cos\theta$ separately. In these plots the statistical

Table 3. The Probability Density $\rho(t/t_{\max}, \cos\theta)$ for τ^+ Decay, obtained from the Empirical Data on 495 Events reported by the Groups listed on p. 196 (excluding Baldo-Ceolin *et al.*) by dividing the Observed Number of Events by the Phase Space Estimate. The Bracketed Numbers give the Actual Number of Events observed in these Data for each Interval

$t/t_{\max}$ (negative pion)	cos θ 0–0·1	0·1–0·2	0·2–0·3	0·3–0·4	0·4–0·5	0·5–0·6	0·6–0·7	0·7–0·8	0·8–0·9	0·9–1·0
0·9–1·0	1·70 (5)	1·01 (3)	2·38 (7)	1·70 (5)	0·34 (1)	0·68 (2)	0·68 (2)	0·68 (2)	1·70 (5)	1·36 (4)
0·8–0·9	1·00 (5)	1·00 (5)	2·00 (10)	0·80 (4)	1·20 (6)	0·20 (1)	1·60 (8)	1·40 (7)	0·40 (2)	1·20 (6)
0·7–0·8	2·12 (13)	0·51 (3)	1·19 (7)	1·02 (6)	1·19 (7)	1·19 (7)	1·19 (7)	1·53 (9)	0·34 (2)	1·02 (6)
0·6–0·7	1·11 (7)	1·43 (9)	1·59 (10)	0·80 (5)	1·11 (7)	0·95 (6)	0·64 (4)	0·80 (5)	0·48 (3)	0·32 (2)
0·5–0·6	1·27 (8)	1·27 (8)	1·11 (7)	0·79 (5)	1·27 (8)	0·95 (6)	1·74 (11)	1·58 (10)	0·95 (6)	0·95 (6)
0·4–0·5	0·49 (3)	0·98 (6)	0·98 (6)	0·65 (4)	0·98 (6)	0·82 (5)	0·82 (5)	1·47 (9)	0·49 (3)	1·19 (7)
0·3–0·4	1·94 (11)	1·05 (6)	0·87 (5)	0·53 (3)	1·58 (9)	1·23 (7)	0·87 (5)	1·05 (6)	0·53 (3)	1·05 (6)
0·2–0·3	1·00 (5)	1·00 (5)	0·40 (2)	1·00 (5)	0·80 (4)	1·20 (6)	1·20 (6)	0·80 (4)	1·20 (6)	1·20 (6)
0·1–0·2	0·75 (3)	0·00 (0)	0·50 (2)	0·25 (1)	1·00 (4)	0·50 (2)	0·25 (1)	0·25 (1)	0·75 (3)	0·50 (2)
0–0·1	1·57 (7)		1·34 (6)		0·67 (3)		0·67 (3)		0·67 (3)	

fluctuations are now relatively small, and the empirical histograms are not consistent with distributions vanishing at $\cos\theta = 1$, at $t = 0$ or at $t = t_{\max}$, unless the distributions vary extremely rapidly in these regions. The rigorous statements apply only at these limiting points, and some estimate of what is a reasonable variation for the probability density in the neighbourhood of these limiting values, for various τ spin-parities, is now desirable.

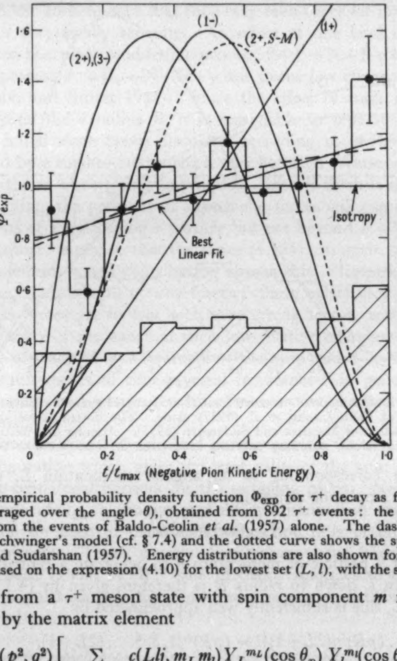


Figure 10. The empirical probability density function $\Phi_{\exp}$ for τ^+ decay as function of π^- kinetic energy (averaged over the angle θ), obtained from 892 τ^+ events: the hatched histogram is obtained from the events of Baldo-Ceolin *et al.* (1957) alone. The dashed curve is the prediction of Schwinger's model (cf. § 7.4) and the dotted curve shows the spin 2+ distribution of Marshak and Sudarshan (1957). Energy distributions are also shown for various spin values; these are based on the expression (4.10) for the lowest set (L, l), with the simple estimate (4.13).

The decay from a τ^+ meson state with spin component m may be described quite generally by the matrix element

$$S(m) = \sum_{Ll} d_{Ll}(p^2, q^2)\left[\sum_{m_L+m_l=m} c(Llj, m_L m_l)\, Y_L^{m_L}(\cos\theta_p)\, Y_l^{m_l}(\cos\theta_q)\right] \quad \ldots(4.10a)$$

In (4.10*a*), the $c(Llj, m_L m_l)$ are the (Wigner) coefficients (cf. Rose 1955) defined by the property that the expression within the bracket corresponds to a total angular momentum j for the pion system, and θ_p, θ_q are the angles made by **p**, **q** with the axis of quantization. The scalar coefficients d_{Ll} are functions only of p, since q is given by equation (4.5). If the axis of quantization is chosen along the direction **p** of the outgoing π^- meson, so that $\cos\theta_p = 1$ and $\cos\theta_q = \cos\theta$, then (4.10*a*) simplifies to

$$S(m) = \sum_{Ll} d_{Ll}[c(Llj, 0m)\, Y_l^m(\cos\theta)], \quad \ldots\ldots(4.10b)$$

since $Y_L^{m_L}(1)$ is zero unless $m_L = 0$, in which case its value is unity.

If a τ meson radius R is defined, outside which the pions may be regarded as moving freely, then, for sufficiently small R, it is reasonable to expect that the momentum dependence of d_{Ll} should be due primarily to the momentum dependence of the penetrability of the centrifugal barriers faced by the outgoing pions, due to their angular momenta L, l. For a particle of momentum p influenced only by the centrifugal barrier corresponding to its angular momentum L, the exact wave function with unit amplitude at infinity is

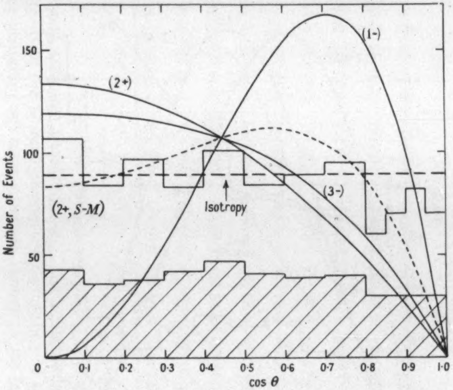


Figure 11. The distribution in cos θ observed for 892 τ^+ decay events—the hatched area corresponds to the events of Baldo-Ceolin *et al.* (1957) alone. The dotted curve shows the spin 2+ angular distribution of Marshak and Sudarshan (1957). Angular distributions are also shown for various spin values, assuming only the lowest set (L, l) in (4.10) to be effective.

$$J_{L+\frac{1}{2}}(pr)\sqrt{(\pi/2pr)}. \qquad \ldots\ldots(4.11)$$

The amplitude for a particle of momentum p to penetrate a centrifugal barrier for angular momentum L down to radius R is therefore given by (4.11), evaluated at $r = R$; for $pR \lesssim L$, this is sufficiently well approximated by

$$(pR)^L/(2L+1)!! = (pR)^L/1.3.5.\ldots(2L+1), \qquad \ldots\ldots(4.12)$$

whose momentum dependence is independent of R. Applying this to each of the degrees of freedom in τ decay, we may write

$$d_{Ll}(p) = \frac{(pR)^L}{(2L+1)!!}\cdot\frac{(qR)^l}{(2l+1)!!}\cdot a_{Ll}(\omega_p), \qquad \ldots\ldots(4.13)$$

where a_{Ll} will generally be a slowly varying function of the π^- energy, the strong momentum dependences of d_{Ll} being removed by the first factors of (4.13).

The expansion (4.10*b*) is completely general. For reasonable radii $R(\lesssim \hbar/m_\pi c)$, the penetration amplitudes of (4.13) fall rapidly with increasing L and l, especially as there is a jump of at least two units of angular momentum from one set (L, l) to the next; for example, with spin 1^-, the lowest set (L, l) is $(2, 2)$, the next set being $(4, 4)$ with penetration amplitudes smaller by a factor of the order $(pR)^2(qR)^2/2500$. However, it cannot be concluded that these higher configurations are necessarily negligible; if there were strong pion–pion forces whose cross section varied rapidly with energy (owing to a resonant interaction, or to a near resonance at zero energy), a strongly interacting π^+–π^- pair may remain together until relatively large radii are reached, with the consequence that they may each carry off rather high angular momentum when they finally separate. An effect of this kind is already known for nucleons after the pion production process $P+P \rightarrow N+P+\pi^+$, and after the pion absorption process $\pi^- + H_2 \rightarrow N+N+\gamma$ due to the low energy nucleon–nucleon resonance (Watson and Stuart 1951). Since the effect of such a resonance takes place mostly beyond the τ radius R, it is reasonable to separate the process into two stages, the initial stage being discussed according to (4.10*b*), the final stage being represented by a further factor which may be taken (Watson and Stuart 1951) to be essentially the π–π scattering amplitude. Even if there is no rapid variation in the π–π interaction, the presence of strong π–π forces will contribute to increase the effective radius of the decaying τ particle but not beyond $R \sim \hbar/m_\pi c$ (since these forces will have short range), so that the series (4.10*b*) can again be used with only low angular momentum waves contributing appreciably. Empirically very little is known concerning the strength of π–π forces; there exists very little evidence to suggest that these forces are in fact such very strong forces, and none to indicate the existence of any π–π resonance at such low relative energies as are relevant in τ decay. In the absence of π–π interactions the amplitudes a_{Ll} may be shown to be real, from requirements of time-reversal invariance; in general, these amplitudes will be complex, with phases resulting from any interactions effective between the outgoing pions.

For certain combinations of spin and parity, namely 0^-, $1^\pm$, 2^+ and 3^-, there is a unique configuration which gives the least value of the sum $(L+l)$ and therefore the weakest centrifugal barriers for the outgoing pions. Neglecting the weak energy dependence of a_{Ll}, a unique probability distribution corresponds to this least configuration for each of these spin values and it is of interest to compare these with experiment. These unique distributions have been given by Fabri (1954) and by Dalitz (1953, 1954); each of them factorizes to the product of an energy and an angular distribution and these separate distributions have been plotted in figures 10 and 11.

Spin–parity values for which a competing 2π-decay is allowed.

Owing to the importance of the question of whether the τ^+ and $K^+_{\pi 2}$ events could represent competing decay modes of a unique K^+ particle, we consider first those spin–parity possibilities for which both 2π- and 3π-decay are permissible.

The case of spin 1^- can be immediately excluded. Taking the lowest configuration (2, 2) alone, the predicted angular distribution disagrees with the experiments in four important respects, since it vanishes strongly (quadratically) at both $t = 0$ and $t = t_{\max}$ and very strongly ($\cos^2\theta$) also at $\cos\theta = 0$, as well as vanishing linearly at $\cos\theta = 1$. This striking disagreement will persist when all higher configurations (which are of the form $l = L = 2n+2$, with $n \geqslant 0$) are included. These higher terms necessarily vanish even more strongly for both slow and fast mesons.

Also point (iii) above requires that (4.10*b*) vanish generally (i.e. for all values of the d_{Ll}) at $\cos\theta = 1$ for spin 1^-. With only $S(m = 0)$ non-zero at $\cos\theta = 1$, this remark requires that each term of $S(m = 0)$ be separately zero ; this implies that $c(Ll\,1,00) \equiv 0$ for each of the sets (L, l) appropriate here (this may also be verified from the explicit expression for Wigner coefficients). Hence $S(m = 0)$ vanishes for all $\cos\theta$. As a result only the $S(m = \pm 1)$ contribute in (4.9*b*), and these vanish for $\cos\theta = 0$ since $Y_l^{\pm 1}(0) = 0$ for all even l; that is, the probability distribution Φ quite generally vanish at least quadratically at $\cos\theta = 0$.

According to the considerations above and in § 4.2.3, spin 2^+ is really the lowest spin value possible for the θ particle and such that τ and θ particles could be identical. This case is therefore of special interest. The distribution for the lowest configuration $(L, l) = (1, 2)$, shown in figures 10 and 11, disagrees strongly with the data in all essential respects. Marshak and Sudarshan (1957) have investigated the effect of including the next configuration (3, 2). The additional terms go in the direction of improving the agreement with the experimental angular distribution but not sufficiently to provide a satisfactory representation ; still it is possible that the inclusion of further configurations could lead to further improvement in this respect. It is the π^- energy distribution which provides the main difficulty for spin 2^+, since the contributions to this from each (L, l) are additive without interference. Since the two lowest configurations require a d-wave for the π^+–π^+ pair, such a high proportion of events giving a π^- meson very close to the maximum energy as that observed is very difficult to reproduce. The inclusion of higher configurations does not help since these require $l = 4$ or higher for the π^+–π^+ pair and therefore reduce further the proportion of fast π^- mesons predicted. To give the high probability observed for very fast π^- mesons, it would be necessary to postulate a resonance near zero energy in the d-wave for two pions of total isotopic spin $T = 2$. This would require a π^+–π^+ interaction strong beyond reason, whereas very little other evidence exists to show the presence of even a moderate π–π force. Even this extreme assumption would not suffice for a good fit despite the fact that it would remove the strongest discrepancy, for the proportion of slow π^- mesons observed has still to be accounted for. It appears that only very complicated assumptions could account for the present data on the basis of spin 2^+, which makes this a very unlikely possibility.

Spins 3^- and 4^+ may similarly be excluded by the difficulty of accounting for the essentially uniform character of the experimental function Φ_{exp} in its dependence on the negative pion energy. Higher spins will be neglected as physically unreasonable, since their large centrifugal barriers would contribute powerfully in slowing down τ-decay, leaving relatively little to be due to the violation of strangeness conservation involved in this decay. Also, despite their electromagnetic character, the absence of competing radiative processes such as

$$\begin{tikzcd} & \pi^+ + \gamma, \\ \tau^+ \arrow[ur] \arrow[dr] & \\ & \pi^+ + e^+ + e^-, \end{tikzcd}$$

becomes difficult to account for when high spin values are considered for the τ particle, since the centrifugal barriers will be far less effective than in 3π-decay owing to the larger energy release, and since their phase space is greater (Dalitz

1955 a); this already provides a problem for spin values as high as 2, with the present body of statistics. On the above basis then, it is concluded that the evidence on τ decay configurations decisively rules out just those spin–parity values which would allow the τ particle to have a competing 2π-decay mode which could account for the observed $K^+_{\pi 2}$ decays.

Spin–parity values for which 2π-decay is forbidden by parity conservation.

The possibility that the τ particle is pseudoscalar (0^-) is now the simplest conclusion in accord with all the data. This would naturally lead to a more or less uniform distribution function Φ. The present experimental data actually indicates quite an appreciable dependence of Φ on the negative pion energy, as well as some dependence on $\cos\theta$, an adequate representation of the present data being given by

$$\Phi = \Phi(0)\{1+0{\cdot}55(\pm 0{\cdot}2)(t/t_{\max})-0{\cdot}24(\pm 0{\cdot}15)\,P_2(\cos\theta)\}. \quad \ldots\ldots(4.14)$$

This energy dependence could perhaps result from a moderate s-wave attraction between π^+ mesons, or from some special features of the decay mechanism (as suggested by Schwinger (cf. § 7.4)), but its really important feature is that it does not vanish for $t = 0$.

For spin 1^+ the main feature expected for the function Φ is that it should vanish for $t = 0$, as remarked in (i) above. Consequently the observation that quite a high proportion of events have a low π^- energy provides evidence against spin 1^+ for the τ meson. This case against spin 1^+ was already stated strongly by Orear, Harris and Taylor (1956) and has become increasingly strong as further data have accumulated. Since the empirical result (4.14) does have an upward slope, it is possible to make spin 1^+ appear somewhat less unacceptable than it does with the simplest estimates, by invoking a suitable energy dependence for a_{10} in the estimate (4.13) and by the inclusion of higher configurations (Eisenberg, Lomon and Rosendorff 1956). However, the observation that 9/495 negative pions have energy less than 2·4 MeV ($t/t_{\max} = 0{\cdot}05$) still provides a strong discrimination against spin 1^+ unless unreasonably extreme assumptions are made. Similarly the case of spin 3^+ may be rejected, a little less decisively since the relative contribution of two configurations $(3,0)$ and $(1,2)$ which have comparable centrifugal barriers provides an adjustable parameter to allow a somewhat better fit than for 1^+; the probability density can then be made to vanish quite steeply at $t = 0$ without such extreme assumptions.

Finally, for spins 2^- and 4^-, there are several configurations which will generally compete (e.g. for 2^-, these are $(2,0)$ and $(0,2)$). These spin values allow s-wave π^- mesons and s-wave relative motion between the π^+ mesons (although not in the same configuration, of course) and give weak angular distributions non-vanishing at $\cos\theta = 0$ or $\cos\theta = 1$. Consequently, when reasonably weak energy-dependence of the coefficients a_{lL} is permitted, these possibilities are in principle very difficult to distinguish from the case of a pseudoscalar τ particle by means of the internal correlations in τ decay alone.

To sum up, we conclude that the statistical data on τ^+ decay configurations points strongly to spin 0^- as the most natural hypothesis for the τ particle, although spins 2^- and even 4^- are equally acceptable on this basis. Other spin values such as 1^+ and 3^+ can be reconciled with this data only with rather extreme assumptions, and spin–parity combinations for which $K^+_{\pi 2}$ decay would be possible are strongly excluded.

Several qualitative arguments against spins equal to or greater than 2 for the τ particle have been mentioned just above. Teucher, Thirring and Winzeler (1955) have suggested a search for angular correlations in the spatial orientation of τ decay events with respect to the production plane and direction as a possible indication of non-zero spin for the τ^+ particle, since any polarization of the τ^+ particle resulting from the production process will not be lost in the slowing down processes (Wolfenstein 1949) nor after the τ^+ particle has come to rest in the emulsion (Thirring 1956). There are many spatial correlations which may be examined (Orear, Harris and Taylor 1956) but no significant deviations from isotropy have yet been established. Unfortunately the observation of spatially isotropic decay provides only a weak argument against non-zero spins, since the production process may be such as to produce no K particle polarizations.

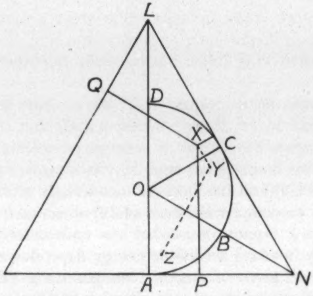


Figure 12. Triangular plot for the representation of τ^+ decay events (see text) (Dalitz 1956 b).

The symmetry between the three pions (irrespective of charge) in τ^+ decay is also of interest (Dalitz 1956 b). The evidence on this may be most readily displayed on a triangular plot (figure 12), where XP, XQ, XR, the perpendiculars from X to the sides of the equilateral triangle LMN, are proportional to the π^- and the π^+ kinetic energies respectively. The requirement that these energies allow momentum conservation confines the points X to the inscribed circle within the triangle, in the non-relativistic approximation; this boundary is considerably distorted when relativistic effects are included (Fabri 1954). The lines OB, OC divide this allowed region into three equal parts and points Y and Z obtained by reflecting X in OC, and then Y in OB, represent configurations with the same pion energies but a different distribution of charges. The data on 219 events are plotted in this manner in figure 13, due to Orear, Harris and Taylor (1956); the distribution of events clearly has a strong component which is symmetrical between the three pions. Unfortunately it is not possible to conclude generally from this that the τ decay amplitude itself is predominantly symmetrical between the pions, although this is the simplest interpretation with spin 0^- for the τ particle; with this assumption a limit may be placed on the non-symmetrical component, which is of interest in discussing the isotopic spin state of the final pions (cf. § 4.4.1).

A K$^+$ decay giving three non-coplanar pions with a total energy of only 43 Mev has been reported by Daniels and Pal (1954). The most reasonable interpretation of this event is as radiative τ^+ decay,

$$\tau^+ \rightarrow \pi^+ + \pi^- + \pi^+ + \gamma \qquad \ldots\ldots(4.15)$$

giving a photon of 32 Mev, since the missing energy and momentum are equal, within experimental error. For a pseudoscalar particle, the predominant mechanism for radiative decay is the internal bremsstrahlung process (Dalitz 1955 a), the probability for emission of a photon of 32 Mev or more being about 10^{-4} relative to normal τ^+ decay. With a total of about 1000 τ^+ decays recorded, the observation of such an event is not unreasonable at this stage. The probability for emitting a photon exceeding 1 Mev is about 3×10^{-3}; however, the observation of such low energy photons is very difficult with present experimental accuracies, the reported Q values

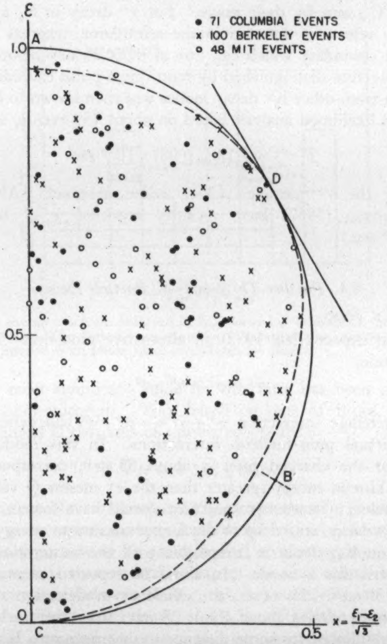


Figure 13. A plot of 219 τ^+ decay events (after Orear, Harris and Taylor 1956). The dashed curve shows the boundary within which events should lie when the relativistic energy-momentum relation is taken into account.

for a batch of 350 τ^+ decays (Baldo-Ceolin *et al.* 1957, and private communication) straggling about ± 2 MeV around their mean value of 75 MeV.

An electronic measurement of the τ^+ lifetime has recently been reported by Motley and Fitch (1957). The counter system used is similar to that used for their measurement of the $K^+_{\mu 2}$ lifetime (cf. figure 9), the important difference being the replacement of counter C_6 by a liquid scintillation counter S_4. The absorber A is chosen to bring the K mesons to rest in S_4, so that the K mesons are separated from protons and pions of the same momentum by their velocity in the Cerenkov counters C_1 and C_2, and by their range. For τ^+ decay in S_4 a large fraction of its 75 MeV energy release is detected in the scintillator, whereas the $K^+_{\pi 2}$ and $K^+_{\mu 2}$ decays give a fast secondary which can lose at most 26 MeV before escaping. The τ^+ decays were therefore distinguished by requiring a pulse exceeding 35 MeV in S_4; the contamination from other K$^+$ decay modes was then shown to be less than 10%. From a maximum likelihood analysis based on about 450 events, a mean lifetime of

$$T(\tau^+) = 1\cdot 17(\pm 0\cdot 08) \times 10^{-8} \text{ sec} \qquad \ldots\ldots(4.16)$$

was deduced for the τ^+ particle. (*Note added in proof*: Alvarez, Crawford, Good and Stevenson (1957) have recently reported a τ^+ lifetime value of $1\cdot 19 \pm 0\cdot 046 \times 10^{-8}$ sec.)

4.4. *Further Three-body* K *Particle Decays*

4.4.1. *The* $K_{\pi 3}$ *or* τ' *decay.*

It is natural to expect (Pais 1952) an alternative pion-decay for the τ meson, the so-called τ' mode,

$$\tau^\pm \rightarrow \pi^\pm + \pi^0 + \pi^0, \qquad \ldots\ldots(4.17)$$

since a charge-exchange interaction $\pi^+ + \pi^- \longleftrightarrow \pi^0 + \pi^0$ will generally be possible through strong virtual pion–nucleon interactions. In this mode, the maximum energy allowed for the charged pion is about 55 MeV, corresponding to a total release of 86 MeV kinetic energy (greater than the τ^+ meson Q value owing to the $\pi^\pm$–π^0 mass difference); its energy spectrum should have form similar to that for the π^- meson in τ^+ decay, scaled up to the higher maximum energy.

The evidence on $K_{\pi 3}$ decay is compatible with the assumption that almost all $K_{\pi 3}$ events represent this τ' mode. In about 60 reported events, there is no π^+ energy exceeding 50 MeV. However, the efficiency of detection and identification falls rapidly for pion energies above about 25 MeV, since the track of a relativistic secondary must be followed for some distance to discriminate a $K_{\pi 3}$ event from the more frequent $K_{\mu 2}$ and $K_{\pi 2}$ events; a slow secondary is detected efficiently and may be quickly identified. No systematic investigation of $K_{\pi 3}$ decay free from this bias has yet been reported. The most extensive investigation is that of Whitehead, Stork, Peterson, Perkins and Birge (1956), whose spectrum based on 34 observed events is given in figure 14. In a selected group of relativisitic secondaries followed to rest in this work, there were identified several $K_{\pi 3}$ events previously missed; these are included in figure 13 with correspondingly large weight. It will be seen that the spectrum then obtained is not in disagreement with expectation for the τ' mode.

In τ' decay, an electron pair resulting from a prompt decay $\pi^0 \rightarrow \gamma + e^+ + e^-$ should be associated with the π^+ meson in about 1/40 cases. One example of this situation has been reported by Heckman, Smith and Barkas (1956 b), compared with a total of about 60 $K_{\pi 3}$ events reported by various groups.

Isotopic spin is clearly not conserved in the τ decay interaction which leads from the $T = \frac{1}{2}$ τ particle to final pions whose total isotopic spin is necessarily integral. An indication of the isotopic spin state for the pions will be given by the τ'/τ ratio and this will be informative concerning the mechanism of the τ decay.

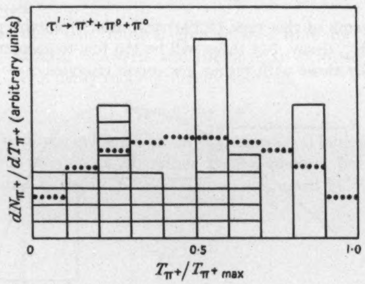


Figure 14. The π^+ energy spectrum observed in $K^+_{\pi 3}$ decay by Whitehead, Stork, Peterson, Perkins and Birge (1956). The dotted lines show for comparison the number of events expected in each energy interval from phase space considerations alone.

The statistically most significant value for this ratio has been given by Birge and Stork (private communication, 1956), based on data of Birge, Perkins, Peterson, Stork and Whitehead (1956) and of associated groups:

$$\tau'/\tau = 0{\cdot}332 \pm 0{\cdot}067. \qquad \ldots\ldots(4.18)$$

Since the phase space for τ' decay is 30% greater than for τ decay (Dalitz 1956 a), the ratio of the mean square matrix elements for τ' and τ decay is 0·255 (±0·05).

Assuming the τ particle to be pseudoscalar, so that s-wave emission is possible for all three pions, it is then natural to expect the final orbital motion of the three pions to be symmetrical between them; this is in accord with the degree of symmetry observed for the normal τ decay (cf. § 4.3). Since the pions are Bose particles, their final isotopic spin state must then also be symmetrical, which is only possible for a total $T = 1$ or $T = 3$ (Dalitz 1956). For the symmetrical $T = 1$ state, a value of $\frac{1}{4}$ is to be expected (Dalitz 1953, Berestetski 1953) for the ratio of the mean square matrix elements for τ' and τ decay; however, this ratio is also possible with other mixtures of $T = 1$ and $T = 3$ states. The simplest explanation is that $T = 1$ for the final pions. This would be in accord with a rule proposed by Gell-Mann (1955) that slow decays should lead to a change $\Delta T = \frac{1}{2}$ in the total isotopic spin of the system (note that $T = 0$ is not possible for final pions of charge $+1$); this interpretation has also been noted by Wentzel (1956 b), by Gatto (1956 a) and by Kawaguchi and Nishijima (1956 a).

4.4.2. *The* $K_{\mu 3}$ *particle.*

Integral spin is known for the $K_{\mu 3}$ particle from the observation of $K_{\mu 3}$ decay following the elementary interaction $\pi^- + P \rightarrow \Sigma^- + K^+$ (Budde, Chretien, Leitner, Samios, Schwartz and Steinberger 1956). With known particles, this implies the emission of a neutrino in $K_{\mu 3}$ decay. The only three-body processes which could give rise to this type of decay are then

$$K^+_{\mu 3} \rightarrow \mu^+ + \nu + \gamma, \qquad \ldots\ldots(4.19a)$$

$$K^+_{\mu 3} \rightarrow \mu^+ + \nu + \pi^0. \qquad \ldots\ldots(4.19b)$$

Of course, some events of this type (4.19*a*) will arise from the process of internal bremsstrahlung in $K^+_{\mu 2}$ decay, but these will be too few to account for the observed $K^+_{\mu 3}$ events, especially those with rather low muon energies.

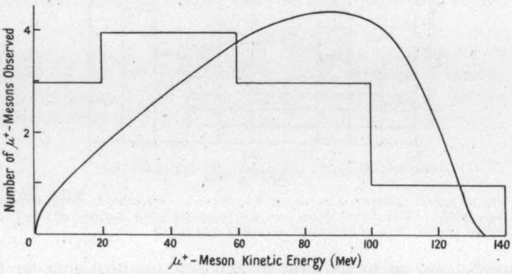


Figure 15. The μ^+ energy spectrum observed in a batch of 19 unbiased $K^+_{\mu 3}$ decay events (Crussard 1956). The curve shown is a theoretical spectrum for emission of an s-wave pion, with a scalar (or pseudoscalar) $\mu + \nu$ coupling.

With the process (4.19*b*), an electron pair resulting from the decay $\pi^0 \rightarrow \gamma + e^+ + e^-$ may be expected to accompany the muon with relative frequency 1/80. Among a total of about 60 recorded $K_{\mu 3}$ decays, two K decay events (Yekutieli, Kaplon and Hoang 1956, Hoang, Kaplon and Yekutieli 1956 a) give a muon with an electron pair, with momenta such that this interpretation is permitted. These momenta do not permit the interpretation $K \rightarrow \mu^+ + \nu + e^+ + e^-$ so that these events cannot be attributed to internal conversion for the photon in the process (4.19*a*). Since the mode (4.19*b*) allows a natural interpretation for these events, which would otherwise have to be interpreted as rare modes involving an additional electron pair alternative to the mode (4.19*a*), it has generally been assumed that the majority of $K_{\mu 3}$ events represent the latter mode. It should be added that the possibility $K^+ \rightarrow \mu^+ + \nu + \pi^0 + \pi^0$ is excluded for a number of clear $K_{\mu 3}$ events (e.g. Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard 1956 a) giving muon secondaries of more than 78 MEV, the maximum muon energy for this four-body decay.

The muon energy spectrum for $K_{\mu 3}$ decay is given in figure 15 for 19 unbiased events obtained from the systematic following of K secondaries. In general, a correction to the observed spectrum will be necessary to allow for the variation of scanning efficiency with muon energy; after making these corrections to the complete $K_{\mu3}$ data of 60 events (Whitehead, Stork, Peterson, Perkins and Birge 1956, Crussard 1956), the final energy spectrum is in good accord with that of figure 15. Unfortunately, the muon spectrum alone is relatively insensitive to the features of the decay interaction which relate most directly to the $K_{\mu3}$ spin. However, the spectrum observed is certainly permissible for a $K_{\mu3}$ particle of spin 0 (either parity); comparison of the observations is made in figure 15 with the spectrum corresponding to s-wave pion emission with a scalar (or pseudoscalar) $(\mu+\nu)$ coupling.

4.4.3. *The* K_{e3} *particle.*

For the K_{e3} particle also, integral spin follows from its observation (Budde, Chretien, Leitner, Samios, Schwartz and Steinberger 1956) following the reaction $\pi^{-}+P\rightarrow\Sigma^{-}+K^{+}$. The decay mode generally assumed,

$$K^{+}_{e3}\rightarrow e^{+}+\nu+\pi^{0}, \qquad \ldots\ldots(4.20)$$

does not disagree with the evidence on 26 K_{e3} decays. There is no direct evidence indicating a neutral pion in (4.20) rather than a photon, but the existence of a corresponding neutral K decay, $K^{0}\rightarrow e^{\pm}+\pi^{\mp}+\nu$ discussed below (§ 5.3) is certainly suggestive.

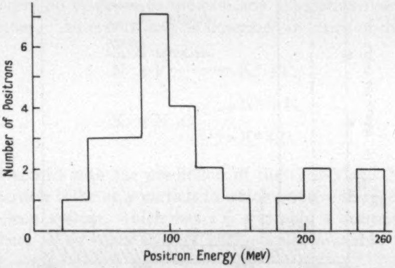


Figure 16. The positron energy spectrum observed in 26 K^{+}_{e3} decay events (Crussard 1956).

The positron energy spectrum observed for K^{+}_{e3} decay is given in figure 16. No correction has been made for the presence of observational bias against low-energy ($\lesssim 30$ MEV) and high-energy electrons. Secondary electrons of low energy may often be lost among the background of slow electrons. For the high energies, the electron can be distinguished from pion or muon only when a relatively long track is available, along which bremsstrahlung energy-loss processes may be observed.

Two of these events give electron energies barely compatible with the upper limit (228 MEV) allowed for process (4.20); for example, Kaplon, Klarmann and Yekutieli (1955) report an event with electron energy 261 ± 34 MEV. It is conceivable that an infrequent $K^{+}_{e2}\rightarrow e^{+}+\nu$ decay mode may exist and be included among this group of K^{+}_{e3} decays. Certainly the relative absence of K^{+}_{e2} decay compared with the K^{+}_{e3} events is clear and very striking in its contrast with the ratio $K^{+}_{\mu2}/K^{+}_{\mu3}\sim 20$ (see table 4). This K^{+}_{e2} absence will undoubtedly form an important clue in our final understanding of beta-decay processes and invites comparison with the rarity ($< 5 \times 10^{-5}$) of $\pi^+ \rightarrow e^+ + \nu$ decay relative to the normal $\pi^+ \rightarrow \mu^+ + \nu$ decay.

4.5. *Nuclear Interactions of* K⁺ *Particles*

Interactions of K⁺ particles with emulsion nuclei have been studied (Lanutti, Chupp, Goldhaber, Goldhaber, Helmy, Iloff, Pevsner and Ritson 1956, Biswas, Ceccarelli-Fabbrichesi, Ceccarelli, Cresti, Gottstein, Varshneya and Waloschek 1956, Biswas, Ceccarelli-Fabrichesi, Ceccarelli, Gottstein, Varshneya and Waloschek 1957, Cocconi, Puppi, Quareni and Stanghellini 1956, Dallaporta 1956) by the systematic following of K⁺ tracks after exposure of the emulsion to the secondary beam. In this work, the initial K⁺ energy has generally been in the range 120–160 MEV and interactions have been observed at various intermediate energies during the slowing down of the K particle. These interactions may be grouped into the following categories :

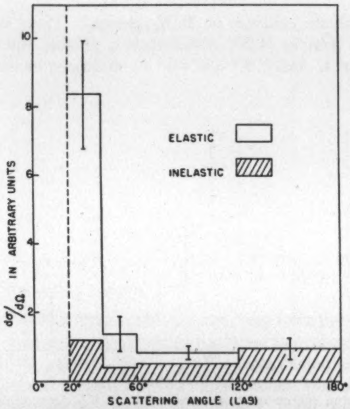


Figure 17. The differential scattering cross section (in arbitrary units) of K⁺ particles by emulsion nuclei (after Lanutti, Chupp, Goldhaber, Goldhaber, Helmy, Iloff, Pevsner and Ritson 1956).

(i) *Elastic* K⁺ *scattering events*, where the K⁺ particle emerges without excitation of the struck nucleus. In collisions with light emulsion nuclei, the recoil track may sometimes be sufficiently long to be observed and to allow a direct check on the energy and momentum balance ; in particular K⁺ collisicns with the hydrogen in the emulsion (see below) may also be identified in this way. In general, however, it has been customary to classify as elastic any K⁺ scattering event without measurable energy loss and without visible nuclear excitation, although it is actually very difficult to establish an energy loss of less than about 5 MEV from the K particle measurements alone.

These elastic events are predominantly due to Coulomb scattering for angles of deflection less than about 30°. However, large angle scatters are observed more frequently than expected for Rutherford scattering, especially when allowance is made for the finite nuclear size, and give evidence for an additional nuclear scattering of the K^+ particle.

(ii) *Inelastic* K^+ *scattering events*, where the K particle emerges with reduced energy, generally resulting in a small star. The angular distribution observed by Lanutti, Chupp, Goldhaber, Goldhaber, Helmy, Iloff, Pevsner and Ritson (1956) for K^+ scattering in emulsion is shown in figure 17, both for elastic and inelastic scattering. In figure 18 the distribution of energy losses for inelastic scattering into the forward and backward hemispheres is given by the histograms F1 and B1.

(iii) K^+ *stops in flight*. The remaining events consist of interactions from which no K particle is seen to emerge. The visible interactions consist of at most a small star; in no case has a pion or hyperon been observed to result and, further, the visible energy of these stars has always been less than the kinetic energy of the incident K particle. With the existence of a neutral K particle known, it is natural to attribute these events to the *charge-exchange scattering* $K^+ + N \rightarrow K^0 + P$. In many of the charge-exchange events there are no visible prongs reported. Some of these events may represent decay-in-flight in which a relativistic secondary has been missed through unfavourable conditions and it is usual to make an empirical efficiency correction to take account of this possibility. The corrected number of decays-in-flight has generally been in excellent accord with the known K^+ lifetimes.

Thus there exists no evidence to indicate any absorptive reaction between K^+ particle and nucleons. All events may be described in terms of the basic scattering interactions

$$K^+ + P \longrightarrow K^+ + P, \qquad \ldots\ldots(4.21a)$$

$$\begin{tikzcd}[row sep=small] & K^+ + N, \qquad \ldots\ldots(4.21b) \\ K^+ + N \arrow[ur] \arrow[dr] & \\ & K^0 + P. \qquad \ldots\ldots(4.21c) \end{tikzcd}$$

This is in direct accord with the prediction of the Gell-Mann–Nishijima scheme, in which the K particle is the only particle to which positive strangeness s is assigned. For the K^+ nucleon system, which has $s = +1$, only a scattering interaction is possible since there is no other system energetically available which includes a baryon and has $s = +1$.

In terms of charge independence, the K^+P interaction (4.21*a*) is characteristic of a $T = 1$ system, since the system has isotopic spin component $T_3 = +1$. The K^+P scattering amplitude will therefore be denoted by a_1 and its total cross section by $\sigma_1 = 4\pi\overline{a_1^2}$. The K^+N system has $T_3 = 0$ and is equally in a $T = 0$ and a $T = 1$ state; replacing the first nucleon in expression (2.3) by a K particle, this initial state may be written $[\phi_{0,0}(K, \mathcal{N}) + \phi_{1,0}(K, \mathcal{N})]/\sqrt{2}$. Denoting the $T = 0$ scattering amplitude by a_0, the state resulting from the K^+–N scattering is

$$[a_0\phi_{0,0}(K, \mathcal{N}) + a_1\phi_{1,0}(K, \mathcal{N})]/\sqrt{2} = \tfrac{1}{2}(a_0 + a_1)\, u_+(K)\, u_N + \tfrac{1}{2}(a_1 - a_0)\, u_0(K)\, u_P, \qquad \ldots\ldots(4.22)$$

so that the scattering amplitude for process (4.21*b*) is $\frac{1}{2}(a_0 + a_1)$ and for process (4.21*c*), $\frac{1}{2}(a_1 - a_0)$. The total K^+N cross section is therefore

$$4\pi \times \tfrac{1}{2}(\overline{a_0^2} + \overline{a_1^2}) = \tfrac{1}{2}(\sigma_1 + \sigma_0).$$

If the target nucleon for the K$^+$ particle is equally frequently a neutron and a proton (roughly true for emulsion nuclei), then the proportion of scatterings giving charge-exchange will be given by

$$\tfrac{1}{4}(a_1-a_0)^2/\{a_1{}^2+\tfrac{1}{2}(a_1{}^2+a_0{}^2)\}. \qquad \ldots\ldots(4.23)$$

If $T=1$ scattering is dominant, the proportion of charge-exchange scattering should have the value 1/6, whereas the value expected is 1/2 if $T=0$ scattering dominates.

From observations on 58·9 metres of track in the range 40–80 MeV, and 56·9 metres for 80–160 MeV by the Dublin, Bristol and Padua groups, the mean free paths in emulsion for the inelastic scatterings (ii) and (iii) are 1·55($\pm$0·25) metres and 1·16($\pm$0·17) metres for the lower and upper energy ranges respectively. These values are compatible with the overall mean 1·36($\pm$0·13) metres for all the data available in this energy range, so that the present evidence is in accord with an essentially constant inelastic cross section about 20% of geometric. This means that the nucleus appears rather transparent to K$^+$ particles, their mean free path within nuclear matter being about 3×10^{-12} cm ; the majority of inelastic K$^+$ scatterings must result from a single interaction, the proportion of events involving double scattering within a nucleus being of order 10%. The mean K$^+$ nucleon cross section within nuclear matter is then $\bar{\sigma}\sim 4$ mbn, where $\bar{\sigma}$ denotes essentially $\tfrac{1}{2}(\sigma_N+\sigma_P)\times\mathscr{P}$, where σ_N and σ_P are the free K$^+$N and K$^+$P cross sections and $\mathscr{P}$ the correction factor giving the proportion of energetically possible final states which are in fact allowed for the recoil nucleon by the Pauli principle when the collision occurs within nuclear matter. The value $\tfrac{1}{2}(\sigma_N+\sigma_P)$ obtained in this way is finally about 6 mbn.

At present 19 charge-exchange events are included in the total of 116 inelastic events (ii) and (iii) reported by the above groups, a proportion of $1/(6{\cdot}1\pm1{\cdot}4)$. Comparing with expression (4.23), the simplest interpretation of this result is that $a_0\ll a_1$, corresponding to K nucleon scattering predominantly in the $T=1$ state, with $\sigma_0\ll\sigma_1$ over the energy range 40–160 MeV.

The analysis of the angular and energy distribution observed for the inelastic scattering events (ii) has been discussed in terms of the Fermi gas model for the nucleus by Biswas, Ceccarelli-Fabbrichesi, Ceccarelli, Gottstein, Varshneya and Waloschek (1957) and by Cocconi, Puppi, Quareni and Stanghellini (1957). With the assumption of an energy-independent and uniform cross section for the K$^+$ nucleon scattering, this model leads to results differing from the experimental data in two respects : (*a*) the proportion of K$^+$ scatterings predicted for the backward hemisphere is very much less than the proportion observed, and (*b*) the mean energy loss predicted for the K$^+$ scatterings into the backward hemisphere is far greater than that observed empirically, as shown in figure 18. The assumption of a backward peaking in the basic K$^+$–nucleon cross section (in centre-of-mass system) would tend to correct the discrepancy (*a*) but would leave (*b*) unaffected. It has been suggested that, if the nucleus represented a repulsive potential for the incoming K$^+$ particle, the effective energy at which the K$^+$–nucleon collision occurs within the nucleus would be lowered, leading to smaller energy losses and to somewhat more pronounced backward scattering. The assumption of a repulsive well of height 10–15 MeV (additional to the Coulomb repulsion) leads to a rather better fit to the observed energy loss distribution. An average nuclear potential of this magnitude for K^+ mesons is in fact quite compatible with a short-range K^+–nucleon interaction whose strength corresponds to a mean cross section $\frac{1}{2}(\sigma_N + \sigma_P)$ of about 6 mbn. However, the backward K^+ scatterings are still rather fewer than observed and, although the situation is still far from clear, it seems probable that some backward peaking in the K^+–nucleon cross section must also be assumed to give a satisfactory fit to the data with this rather crude model.

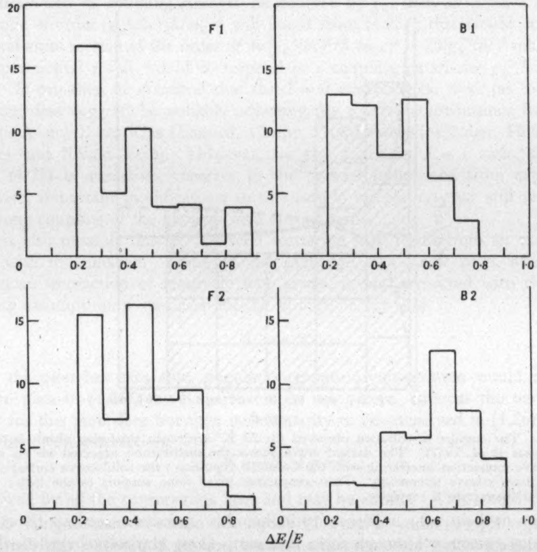


Figure 18. The distribution of proportionate energy loss ΔE/E for K^+ mesons scattered by emulsion nuclei into the forward and backward hemispheres (Cocconi, Puppi, Quareni and Stanghellini 1956). The experimental data available at present is shown in F1 and B1, and is to be compared with the histograms F2 and B2 calculated with a constant, energy independent K^+–nucleon cross section and a Fermi gas model for the nucleus.

Osborne (1956) has pointed out that, as a result of the weakness of K^+–nucleon scattering, the interference between Coulomb and elastic nuclear K^+ scattering will occur for angles of 20–30° where its examination is not difficult. His preliminary analysis of the data of Lanutti, Chupp, Goldhaber, Goldhaber, Helmy, Iloff, Pevsner and Ritson (1956) suggested destructive interference, but the more recent data appear to indicate constructive interference appropriate to K^+ scattering by a repulsive nuclear well of about 10 MeV (Cocconi, Puppi, Quareni and Stanghellini 1957), in accord with the analysis of the inelastic scattering. However, in view of the difficulty of distinguishing between elastic and inelastic scattering, this conclusion must still be regarded with some reserve.†

K^+ collisions with free protons in emulsion have been reported by Biswas, Ceccarelli-Fabbrichesi, Ceccarelli, Cresti, Gottstein, Varshneya and Waloschek (1956), who found eleven such events along 124 metres of K^+ track above 40 MEV and by the Padua group (Dallaporta 1956), who found four events in 80 metres. The Bristol–Dublin investigation has reported no K^+–hydrogen collisions in a total of 75 metres of K^+ track. Taken together these data correspond to a total cross section $\sigma_1 = 15(\pm 3)$ mbn. Figure 19 shows the centre-of-mass angular distribution for these events; although some forward peaking is possible, this distribution does not differ significantly from a uniform differential cross section. On the other hand, the cross section σ_1 obtained appears rather large in comparison with the value $\frac{1}{4}(3\sigma_1+\sigma_0)\sim 6$ mbn indicated by the K^+–nucleus scattering, even taking account of the charge-exchange result $\sigma_0 \ll \sigma_1$.

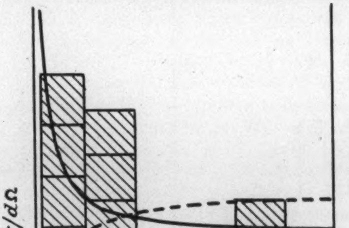


Figure 19. The angular distribution observed for 22 K^+–hydrogen scattering events in emulsion (Biswas *et al.* 1957). The dashed curve shows the distribution expected for an attractive s-wave interaction interfering with the Coulomb repulsion; the solid curve corresponds to a repulsive s-wave interaction. This comparison gives some support to the belief that the $T = 1$ interaction is repulsive.

At present the theoretical interpretation of these data on the K^+–nucleon interaction is not at all clear. With the known interaction

$$\mathcal{N} \longleftrightarrow \mathrm{K} + \begin{pmatrix}\Lambda\\ \Sigma\end{pmatrix}, \qquad \ldots\ldots(4.24)$$

---

† *Note added in proof*: Angular distribution curves shown by Dallaporta (1957) of elastic K^+ scattering by emulsion nuclei at small angles are very closely fitted by the assumption of constructive interference between nuclear and Coulomb scattering and strongly disagree with the curve for destructive interference. Ceccarelli (private communication) has also pointed out that the existence of an attractive potential sufficiently strong to give this result would require that the formation of a K^+–nuclear fragment should occur in quite a high proportion of the inelastic collisions, contrary to observation.

a K^+–nucleon scattering interaction can be built up from sequences of virtual processes (4.24), for example

$$K^+ + P \rightarrow K^+ + \begin{pmatrix}\Lambda^0\\ \Sigma^0\end{pmatrix} + K^+ \rightarrow P + K^+. \qquad \ldots\ldots(4.25)$$

In these intermediate states there is necessarily a large energy unbalance $\Delta E \gtrsim m_K c^2$ so that the resulting interaction is of short range $\hbar c/\Delta E \lesssim \hbar/m_K c$, according to the uncertainty principle. If the interaction (4.24) is assumed to be strong for S-wave K particles and its coupling strength be denoted by g_K, then an S-wave scattering amplitude of order $(g_K^2/\hbar c)(\hbar/m_K c)$ will result from (4.25); this would correspond to a total cross section of the order of $4\pi(g_K^2/\hbar c)^2(\hbar/m_K c)^2 = 20(g_K^2/\hbar c)^2$ mbn, so that the experimental result would correspond to a coupling parameter $g_K^2/\hbar c$ of order unity. It can then be arranged that the $T = 0$ scattering be weak (as the charge-exchange data suggest) by suitably adjusting the relative contributions from intermediate Λ and Σ particles (Lanutti, Chupp, Goldhaber, Goldhaber, Helmy, Iloff, Pevsner and Ritson 1956). However, for the dominant $T = 1$ state, this interaction (4.25) is attractive, contrary to the present indications from experiment; of course, important modifications to this simple calculation may still result from the strong coupling of the nucleon with virtual pions.

It is also possible that K^+–nucleon scattering may result from an exchange of pions with the nucleon. However, the exchange of a single pion, which would lead to an interaction of relatively long range, is not permitted with the natural working assumption of spin zero for the K particle, for then

$$K \not\leftrightarrow K + \pi. \qquad \ldots\ldots(4.26)$$

Since the pion has zero spin, angular momentum conservation would require an S-wave pion in (4.26), the K particle spin being zero, whereas the odd intrinsic parity for this pion does not then permit parity to be conserved in (4.26). On the other hand, the interaction

$$K \longleftrightarrow K + \pi + \pi \qquad \ldots\ldots(4.27)$$

is allowed by all the conservation laws and may be strong. This would also contribute to K–nucleon scattering, although the interaction of a nucleon with a pion pair is believed relatively weak from the evidence on low-energy pion–nucleon scattering and on other processes involving a pair of pions.

However, the situation is complicated further by the existence of at least two distinct K particles, the τ and θ, with spin–parity 0^- and 0^+ respectively. *A priori*, it would be natural to expect quite different nuclear scattering for these two particles; if the interaction (4.24) leads to strong S-wave scattering for one of these K particles, then strong scattering from this interaction would be possible only in the p-wave for the other K particle. In fact, as will be discussed in § 4.6, the relative frequencies of K^+ decay modes are essentially unaffected by nuclear scattering of the K^+ beam, which requires that the cross sections for τ^+ and θ^+ nuclear scattering be rather similar over the energy range 40–160 MEV. One possible interpretation of the equality of τ^+ and θ^+ cross sections involves the 'parity-conjugation' hypothesis of Lee and Yang (1956 a) according to which two sets of Λ and Σ hyperons with opposite parities must exist, the Λ_1, Σ_1 set having an interaction (4.24) with the θ particle of exactly the same form as the interaction (4.24) of the Λ_2, Σ_2 set with the τ particle; this will be discussed further in § 7.3.

The existence of two K particles of opposite parity now allows the possibility of a strong pion interaction

$$\tau \longleftrightarrow \theta + \pi, \qquad \ldots\ldots(4.28)$$

since the emission of an S-wave pion is now consistent with parity conservation, as well as with strangeness conservation and charge independence. This interaction (4.28) would give rise to a relatively long-range interaction ($\sim \hbar/m_\pi c$) for a scattering process such as

$$\tau^+ + \mathrm{P} \rightarrow \theta^+ + \pi^0 + \mathrm{P} \rightarrow \theta^+ + \mathrm{P} \qquad \ldots\ldots(4.29)$$

in which τ particle is transformed to θ particle, or vice versa. This interaction cannot give rise to scattering without energy loss from any target nucleus of zero spin (e.g. $\tau^+ + {}^{12}\mathrm{C} \not\rightarrow {}^{12}\mathrm{C} + \theta^+$ is forbidden since angular momentum conservation is incompatible with parity conservation for this process). Also, in all cases, this scattering is incoherent with Coulomb scattering, since the latter does not change the K particle parity.

Owing to its relatively long range, the scattering process (4.29) could account for the total K$^+$P cross section with quite a moderate coupling strength ${g_{\pi\mathrm{K}}}^2/{m_\pi}^2 \simeq 0{\cdot}2$, for the interaction (4.28). However, since the momentum transfers effective in K$^+$–nucleus scattering in the present energy range already exceed $m_\pi c$, the long range of this interaction would correspond to a cross section falling rapidly with increase in energy or angle of scattering, and this feature is not apparent in the K$^+$ scattering data at present.

The main features of the K$^+$ scattering data are consistent with the existence of a strong, short-range repulsion in the $T = 1$ state, the $T = 0$ interaction being relatively weak. This situation does not have a simple correspondence with any of the mechanisms discussed above, and no clear theoretical interpretation of the data is yet available. However, there are many mechanisms which may all contribute to the K–nucleon interaction, and these are complicated by the existence of two K particles, the τ and θ. More accurate information on the energy and angular dependence of the elementary K–nucleon scattering processes will certainly be desirable for guidance between these many possibilities.

4.6. *The Relative Frequencies of* K$^+$ *Decay Modes*

The most detailed determination of relative K$^+$ frequencies has been that of Birge, Perkins, Peterson, Stork and Whitehead (1956), for the Berkeley secondary beam obtained for circulating protons of 6·2 GeV. In this work the K$^+$ particles had 350 MeV/c momentum and reached the emulsion after a proper flight-time of $1{\cdot}4 \times 10^{-8}$ sec. Clearly attention must be given to assessing the relative efficiencies with which these various decay modes may be identified, as was necessary for determining the energy spectrum for the three-body decay modes. The final results of Birge, Perkins, Peterson, Stork and Whitehead are given in the first column of table 4. The $\mathrm{K}_{\mu 2}$ and $\mathrm{K}_{\pi 2}$ particles form the major part of the secondary beam, amounting to about 60% and 30% respectively. The τ and τ' modes together account for about 7% of the K particles, the proportion $\tau'/(\tau + \tau')$ being close to 1/4.

The remaining three-body decay modes appear still less frequently,† the $K_{\mu 3}$ forming 2·6% of the K^+ decays, the K_{e3} mode about 2%.

It is of particular interest to compare these results with relative frequencies observed for K particles in other situations. If there are several distinct K particles with different spin and parity values these would generally be expected to have different lifetimes, different production characteristics, different cross sections for nuclear scattering, and so on. In this way one may hope to find evidence throwing light on the relationships between the various K^+ modes.

Relative K^+ frequencies have also been reported for secondary beams corresponding to lower proton energies or different angles of emission. At Berkeley, Smith, Heckman and Barkas (1956) have established these ratios for the 90° secondary beam (375 MEV/*c*) for 4·8 GEV incident protons. At Brookhaven, Hoang, Kaplon and Yekutieli (1956 b) have exposed emulsions to a secondary beam (270 MEV/*c*) emitted at 60° to the 2·9 GEV incident proton beam. Their relative frequencies for various K^+ modes are listed in table 4, and do not differ, within statistics, from those of Birge, Perkins, Peterson, Stork and Whitehead.

Table 4. Relative Frequencies observed for K^+ Decay Modes

B \ A	(1)	(2)	(3) (*a*)	(3) (*b*)	(4)	(5)	(6)	(7)
$K_{\mu 2}$	57·2 ± 5·0	65·2 ± 10	60·1 ± 12	61·5 ± 15	63 ± 8	54 ± 6	55 ± 20	60 ± 9
$K_{\pi 2}$	31·2 ± 3·5	21·4 ± 5·6	23·5 ± 6	20·6 ± 6	19 ± 5	26 ± 5	23 ± 8	30 ± 6
$\tau + \tau'$	6·9 ± 0·7	9·7 ± 1·6	7·9 ± 1·7	8·1 ± 2	(6·7)	8 ± 4	7 ± 2	6·9 ± 1·0
$K_{\mu 3}$	2·6 ± 1·1	1·1 ± 0·7	4·2 ± 1·3	4·0 ± 2	2·5 ± 1	4 ± 2	9 ± 4	2·7 ± 2·0
K_{e3}	2·1 ± 1·5	1·9 ± 0·8	4·7 ± 3·5	5·5 ± 3	9·0 ± 4	8 ± 4	6 ± 6	—

A Experimental conditions.
B Decay mode.
(1) Birge, Perkins, Peterson, Stork and Whitehead (1956): bevatron exposure 6·2 GEV, 350 MEV/*c*, 90°.
(2) Smith, Heckman and Barkas (1956): bevatron exposure 4·8 GEV, 375 MEV/*c*, 90°.
(3) Hoang, Kaplon and Yekutieli (1957): cosmotron exposure 2·9 GEV, 270 MEV/*c*, 60°, flight distance (*a*) 352 cm, (*b*) 427 cm.
(4) G-stack collaboration (1956): cosmic-ray exposure.
(5) Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard (1956 a): emulsion exposure to 6·2 GEV protons.
(6) Biswas, Ceccarelli-Fabrichesi, Ceccarelli, Gottstein, Varshneya and Waloschek (1956 b): emulsion scattered K particles.
(7) Widgoff, Shapiro, Schluter, Ritson, Pevsner and Henri (1956): cosmotron exposure 2·9 GEV, 340 MEV/*c*, 120°.

K particles produced and brought to rest in emulsion generally have a flight-time of order 5×10^{-10} sec, short compared with that for the secondary beams, so that any relatively short-lived K particle decays would appear with correspondingly increased frequency in these circumstances. Relative K^+ frequencies have been reported for locally produced K particles, both for cosmic ray production in the G-stack collaboration (1956) and for production by 6·2 GEV protons by Crussard,

---

† *Note added in proof*: Quite recently O'Ceallaigh and his collaborators (1957) have studied the three-body K^+ decay modes paying special attention to the avoidance of bias against events with energetic secondaries and to their separation from the two-body modes. They have concluded that the three-body modes occur relatively more frequently than shown in table 4, the percentage frequencies being 56·9 ± 2·6, 23·2 ± 2·2, 8·9 ± 0·6, 5·9 ± 1·3 and 5·1 ± 1·3 in the order given in the table.

Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard (1956 a), who exposed a stack to the internal beam at Berkeley. In these investigations the K_{e3} mode had frequency 8–9%, but with such large statistical error that this is not incompatible with the first column.

A direct examination of the effect of nuclear scattering on relative K^+ frequencies has been made by Biswas, Ceccarelli-Fabrichesi, Ceccarelli, Gottstein, Varshneya and Waloschek (1956 b), who compared the decay modes for K particles from the secondary beam which had suffered large-angle scattering in the emulsion with those for the K particles of this beam which came to rest without nuclear scattering. With relatively poor statistics, the important conclusion is that the τ and $K_{\pi 2}$ proportions are not appreciably affected by nuclear scattering (cf. table 4). Widgoff, Shapiro, Schluter, Ritson, Pevsner and Henri (1956) have also reported an experiment bearing indirectly on this same point; K^+ frequencies are determined for a secondary beam (340 MeV/*c*) emitted at 120° to 2·9 GeV protons incident on a uranium target. Now K^+ particles in this beam can result directly from nucleon–nucleon collisions only if the target nucleon has a rather large Fermi momentum. However, K particles produced in forward directions can undergo scattering within the same nucleus and thereby also contribute to this secondary beam; calculations suggest that about 80% of the 120° beam results from the scattering process compared with about 20% for the 60° beam. The relative frequencies determined for this 120° beam (flight time $1{\cdot}4 \times 10^{-8}$ sec) are listed in the table and appear to be essentially unaffected by the nuclear scattering which the majority of these K particles have undergone.

The relative frequencies listed in table 4 do not show any systematic deviations but are all compatible with the mean frequencies. This indicates that, if there are distinct K particles, these must have all essentially the same lifetime (within a factor 2), closely related production processes (the ratios being unchanged from 2·9 to 6·2 GeV) and the same nuclear scattering.

4.7. *The Nuclear Interaction and Decay of* K^- *Particles*

Despite the low K^- intensities available in the accelerator beams up to the present (cf. § 3.3), studies of K^- particle interactions (both in flight and at the end of their range) with emulsion nuclei have been reported by a considerable number of groups (Chupp, Goldhaber, Goldhaber and Webb 1955, Fry, Schneps, Snow and Swami 1955 a, b, George, Herz, Noon and Solntseff 1956, Hornbostel and Salant 1956, Haskin, Bowen and Schein 1956, White *et al.*, quoted by S. Goldhaber 1956, Fournet and Widgoff 1956). These data have recently been summarized and discussed by Goldhaber (1956). Further, a direct investigation of the elementary K^-–proton interaction has recently become possible (Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp 1956) with the use of a 10-inch hydrogen bubble chamber at Berkeley.

Along about 17 metres of K^- tracks followed in emulsion, a total of 58 nuclear interactions in flight have been reported by the above groups; this corresponds to a mean free path of $30{\cdot}5(\pm 4)$ cm, which means essentially geometric cross section for the interaction of K^- particles with nuclei. These interactions generally produce quite large stars in which the visible energy released exceeds the K^- kinetic energy (<100 MeV); the prong distribution observed is shown in figure 20(*b*).

Further, the emission of a charged Σ particle is observed from about 15% of these stars, the emission of a charged pion from about 30%. Clearly strong absorptive reactions predominate in the interaction of K^- particles with nuclei. A few events representing elastic nuclear scattering of K^- particles have been reported by various groups (e.g. Hornbostel and Salant 1956) and undoubtedly some events may represent charge-exchange scattering of the K^- particle although this possibility can definitely be excluded for a high proportion of the events. The nuclear interactions of the K^- particles are therefore qualitatively different from those of the K^+ particles, discussed in § 4.5.

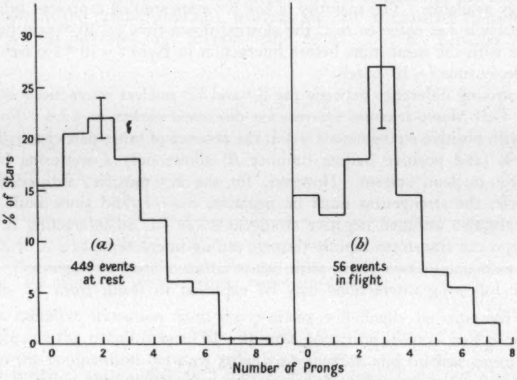


Figure 20. The prong distribution observed for K^- capture stars for K^- particles (*a*) after coming to rest, and (*b*) in flight (*after* Goldhaber 1956).

Three events which are dynamically consistent with K^- elastic scattering from hydrogen in the emulsion have recently been described by Gilbert, Violet and White (1956 b). The centre-of-mass scattering angles for each of these events were 112° (at K^- energy 35 MEV), 22° (at 107 MEV) and 135° (at 32 MEV). Taken together with the work mentioned in the last paragraph, these events correspond to a cross section of 44(± 25) mbn for $K^- + P$ elastic nuclear scattering.

Decays-in-flight for K^- particles have also been observed in this K^- track scanning. On the basis of 13 decay events, Iloff, Goldhaber, Goldhaber, Lanutti, Gilbert, Violet and Widgoff (1956) have deduced a lifetime (averaged over all the decay modes)

$$T(K^-) = 0{\cdot}95^{+0{\cdot}36}_{-0{\cdot}25} \times 10^{-8} \text{ sec} \qquad \ldots\ldots(4.30)$$

which is compatible with the known K^+ lifetimes. Two of these K^- decays have been shown to be consistent with $K^-_{\pi 2}$ and K^-_{e3} modes (Ekspong and Goldhaber 1956). Clear examples of τ^- decay have been known for some time from cloud chamber work (e.g. Van Lint and Trilling 1953).

K^- particles which reach the end of their range will be captured into atomic orbits within the atoms of the emulsion. These K^- particles will then cascade down through successive atomic orbits, first by ejecting Auger electrons from the atom and later by making radiative transitions, until a low-lying orbit is reached from which nuclear capture takes place. Auger electrons with sufficiently high energy (>15 kev) may readily be detected in emulsion and are in fact observed for about 20% of these K^- stars (Fry, Schneps, Snow and Swami 1955 a, b). These capture stars are found to be quite similar to the stars from interaction in flight, as one would expect for a strongly absorptive K^- interaction as long as the initial K^- kinetic energy is smaller than the energy released in the absorption reaction. Their prong distribution is shown in figure 20(*a*); the average prong number is naturally a little less than for interactions in flight since the star has correspondingly less energy available. The majority of the K^- stars studied represent capture after the K^- particle has come to rest, the slowing-down time ($\lesssim 10^{-10}$ sec) being short compared with the mean time before interaction in flight ($\sim 10^{-9}\, c/v$ sec) and with the K^- decay time ($\sim 10^{-8}$ sec).

This striking difference between the K^+ and K^- nuclear interactions is in accord with the Gell-Mann scheme, for reasons discussed earlier in § 2.4. For the K^+ particle with positive strangeness $s=+1$, the absence of other particles with positive strangeness (and positive baryon number B) allows only a scattering interaction for the K^+–nucleon system. However, for the K^- particle, antiparticle to the K^+ particle, the strangeness must be negative, $s=-1$, and since both Λ and Σ particles are also assigned negative strangeness $s=-1$, an interacting K^- particle and nucleon can transform rapidly through strong interactions to a Λ or Σ hyperon together with one or two pions, with conservation of total strangeness s. On this basis, the following interactions may be expected to result from K^- absorption by a proton:

$$K^- + P \rightarrow \begin{cases} \Sigma^+ + \pi^- + 103 \text{ Mev}, & \ldots\ldots(4.31a) \\ \Sigma^0 + \pi^0,\ \Sigma^0 \rightarrow \Lambda^0 + \gamma, & \ldots\ldots(4.31b) \\ \Sigma^- + \pi^+ + 95 \text{ Mev}, & \ldots\ldots(4.31c) \\ \Lambda^0 + \pi^0 + 182 \text{ Mev}, & \ldots\ldots(4.31d) \\ \Lambda^0 + \pi^+ + \pi^- + 38 \text{ Mev}, & \ldots\ldots(4.31e) \\ \Lambda^0 + \pi^0 + \pi^0 + 47 \text{ Mev}. & \ldots\ldots(4.31f) \end{cases}$$

For K^- absorption by a neutron, the corresponding reactions are

$$K^- + N \rightarrow \begin{cases} \Sigma^- + \pi^0, & \ldots\ldots(4.32a) \\ \Sigma^0 + \pi^-, & \ldots\ldots(4.32b) \\ \Lambda^0 + \pi^-, & \ldots\ldots(4.32c) \\ \Lambda^0 + \pi^- + \pi^0. & \ldots\ldots(4.32d) \end{cases}$$

K^- *capture by protons.* Considerable information on these processes (4.31) has already been obtained (Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp 1956) from exposures of a hydrogen bubble-chamber to the K^- beam, absorbing material being placed in front of the chamber to bring the K^- particles to rest within the chamber, the pions in the beam passing on beyond. The majority of the interactions observed took place for K$^-$ particles which had clearly come to rest, and we shall consider only these interactions here.

The K$^-$–proton capture reaction will take place after the K$^-$ particle has reached a low-lying Bohr orbit of the K$^-$–proton system. In contrast with the case of the π^- meson, for which (π^-P) capture is known to occur almost entirely from s-orbits (Brueckner, Serber and Watson 1951), it is possible that strong absorption of K$^-$ particles may also occur from p-orbits about the proton (Gatto 1956 b). Owing to the decrease ($\sim 1/m$) in the radius of a Bohr orbit with increasing mass m of the particle, the rate of nuclear capture from a p-orbit increases with the fifth power of m (the probability is proportional to $|\text{grad}\,\psi(0)|^2$, the square of the slope of the wave function at the proton position), whereas the rate of radiative transition from p-orbit to s-orbit increases only proportionally to m. The K$^-$ particles which do undergo a radiative transition from the p-orbits will finally be captured from an s-orbit. Hence the proportion of K$^-$ captures occurring from s- and p-orbits will depend on the competition between radiative transition and nuclear capture in the p-orbits. Whether this competition will favour K$^-$ absorption from the p-orbit will depend, naturally, on the specific interactions effective between K particles, nucleons and hyperons, and also on their spin values, and therefore cannot be decided at the present stage.

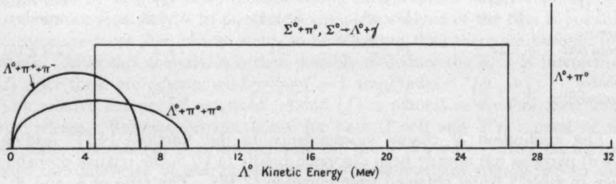


Figure 21. The energy distribution of Λ^0 particles expected for the various processes (4.31) of K$^-$ capture by protons (relative magnitudes arbitrary).

Since a magnetic field is present in the chamber, the processes (*a*), (*c*) and (*e*) of (4.31) will be readily identifiable. Owing to the low velocity of the Λ^0 particles produced in reactions (*b*), (*d*), (*e*) and (*f*) and to the size of the chamber, the observational efficiency for the reactions (4.31) leading to neutral products will be very high whenever $\Lambda^0\rightarrow P+\pi^-$ decay occurs; a correction for the proportion of neutral decays $\Lambda^0\rightarrow N+\pi^0$ could readily be made if only this branching ratio (cf. § 6.1) were known with greater certainty. The Λ^0 particles from reaction (*d*) are mono-energetic (29 Mev) whilst those from reaction (*f*) have at most 9·5 Mev; the observations of Λ^0 particles with kinetic energies lying between these values must be attributed to reaction (*b*) and have therefore provided rather convincing evidence for the existence of the Σ^0 particle. The general form expected for the distribution of Λ^0 energies is shown in figure 21. The uncertainties in measurement make it difficult to distinguish between reactions (*b*) and (*d*) for a Λ^0 particle whose energy is near the upper end of this spectrum: this is the reason for the correlated uncertainties quoted in columns (*b*) and (*d*). The events without visible reaction products, listed in the last column, are presumably due mainly to capture reactions (*b*) and (*d*) in which the Λ^0 particle has undergone a neutral mode of decay. Naturally there is considerable uncertainty in the number of these events. A small proportion of the π^- mesons in the K^- beam may come to rest in the chamber and capture of these will generally not give rise to visible products, also there may be some contribution from charge exchange of K^- mesons near the end of their range since it is difficult to establish whether or not a K^- meson has come to rest before interaction. However, the ratio of columns (*b*) and (*d*) to the total number of K^- interactions giving neutral products, namely $0{\cdot}45 \pm 0{\cdot}10$, is consistent with the values which have been obtained for α_Λ, the proportion of charged Λ^0 decays.

Table 5. The Relative Probabilities observed and calculated for Various K^-–proton Capture Reactions

$K^- + P \rightarrow$	(*a*) $\Sigma^+ + \pi^-$	(*b*) $\Sigma^0 + \pi^0$	(*c*) $\Sigma^- + \pi^+$	(*d*) $\Lambda^0 + \pi^0$	No visible product
Capture from rest	27	14 ± 2	51	7 ∓ 2	$25(\pm 10)$
Relative probability	$\left(\frac{a_0}{\sqrt{6}} + \frac{a_1}{2}\right)^2$	$\frac{a_0^2}{6}\alpha_\Lambda$	$\left(\frac{a_0}{\sqrt{6}} - \frac{a_1}{2}\right)^2$	$\frac{b_1^2}{2}\alpha_\Lambda$	—

The application of charge independence to the reactions (4.31) leading to $(\Sigma + \pi)$ particles has already been discussed briefly in § 2.5, the relative probabilities given in table 5 being obtained from equation (2.32). The ratio of a_0 and a_1, the amplitudes for the $T = 0$ and $T = 1$ transitions from the initial $K^- + P$ system, may be obtained from

$$a_1^2/a_0^2 = [n(\Sigma^+) + n(\Sigma^-) - 2n(\Sigma^0)]/3n(\Sigma^0), \qquad \ldots\ldots(4.33)$$

where $n(\Sigma^+), n(\Sigma^0), n(\Sigma^-)$ denote the relative frequencies for the $\Sigma^+, \Sigma^0, \Sigma^-$ reactions (4.31) after correcting the observed number of Σ^0 reactions by the factor $1/\alpha_\Lambda$, α_Λ being the proportion of Λ^0 particles undergoing $\Lambda^0 \rightarrow P + \pi^-$ decay. With variable allowable estimates for α_Λ (cf. § 7.1), the results quoted in table 5 lead to values of a_1^2/a_0^2 ranging from zero (for $\alpha_\Lambda = 0{\cdot}35$) up to $0{\cdot}6$ (for $\alpha_\Lambda = 0{\cdot}65$). If the reaction proceeded through only the $T = 0$ channel or only the $T = 1$ channel, the ratio Σ^+/Σ^- would be 1. In general, this ratio reflects the relative magnitudes of a_0 and a_1 through an interference between these two channels, and the above results correspond to an interference term which requires a ratio a_1/a_0 no less than about $0{\cdot}13$. If capture proceeds only for K^- mesons in the atomic s-state, then, accepting the value $\alpha_\Lambda = 0{\cdot}65$ (cf. § 7.1), the smallness of this interference term would imply a phase difference of about $\pm 108°$ between the $T = 0$ and $T = 1$ reaction amplitudes. However, if the capture takes place from both s and p states, as seems likely, the relative frequencies expected for the various reactions (4.31) are to be obtained as the average of the relative probabilities of table 5 for each initial $K^- + P$ atomic state, weighted according to the total capture rate from each state; in this case the smallness of the interference term observed could result from cancellations between the interference terms corresponding to each of the initial states, without implying such large phase differences between the various reaction amplitudes.

The $(\Lambda+\pi)$ reaction (4.31*d*) can take place only from the $T=1$ initial state and the amplitude for this transition has been denoted by b_1. The phase space available for the final $(\Lambda+\pi)$ system is proportional to the pion momentum; owing to the greater energy release, the phase space for the $(\Lambda+\pi)$ reaction is about 1·7 times that for the $(\Sigma+\pi)$ reaction. A direct comparison of observed Σ^0 and Λ^0 decays gives the ratio of b_1 to a_0 at once,

$$b_1{}^2/a_0{}^2 = (7\pm2)/3(1{\cdot}7)(14\mp2) = 0{\cdot}10{}^{+}0{\cdot}04 \qquad \ldots\ldots(4.34)$$

The $T=1$ channels, both for Σ and Λ reactions, are therefore somewhat weaker than the $T=0$ channel effective for the Σ reaction alone.

The uncertainties in the physical situation do not allow a clear-cut analysis at the present stage, beyond the above remarks. Observations on the angular distribution of the reaction products for K^- capture in flight as function of incident energy would be of great value in establishing these capture rates for s-wave and p-wave capture separately. In principle a complete analysis of the data is possible, since there are more $(\Sigma\pi)$ charge states to be observed than there are isotopic spin channels. After this analysis, it is then possible to deduce the $\mathrm{K}^-+\mathrm{N}$ interactions (4.32) since these are determined by the $T=1$ amplitudes.

The relative absence of reactions (*e*) and (*f*) is natural in view of their small energy release. Reaction (*e*) can occur for both $T=0$ and $T=1$ parts of the initial state, whereas (*f*) is permitted only for the $T=0$ part. After taking into account the difference of energy release and the symmetries appropriate to each isotopic-spin state, it may be deduced that reaction (*f*) cannot exceed (*e*)—the absence of these readily identifiable events (*f*) therefore makes it reasonable to neglect the reaction (*e*).

The angular correlations observed between the pion direction in the $\Sigma^\pm$ decay and the direction of the Σ motion are given in table 6. This angular distribution is compatible with an isotropic distribution, although there is some suggestion of a dip[†] around 90°. Treiman (1956 a) has remarked that, if the K particle has spin zero and is captured from an s-orbit, this angular distribution depends uniquely on the Σ spin; however, there is considerable doubt whether the capture process occurs only from s-states (Gatto 1956 b). In any case, observation of any significant angular correlation would imply spin greater than $\frac{1}{2}$ for the $\Sigma^\pm$ particles. The energy distribution of Λ^0 decays from reaction (*b*) is also of interest in this respect, since this would reflect the angular correlation in the $\Sigma^0\rightarrow\Lambda^0+\gamma$ decay following reaction (4.31*b*).

Table 6. The Angular Correlation in $\Sigma^\pm$ Decay following K^-–proton Capture[†]

	0–60°	60°–90°	90°–120°	120°–180°
$\Sigma^-\rightarrow\mathrm{N}+\pi^-$	12	10	6	13
$\Sigma^+ \nearrow \mathrm{P}+\pi^0$	5	4	0	3
$\searrow \mathrm{N}+\pi^+$	4	2	2	3
Totals	21	16	8	19

[†] *Note added in proof*: More recent data (Rochester Conference 1957) show no evidence for a dip around 90°, but correspond to completely isotopic Σ decay. Of 53 Σ^+ decays, 25 events gave a pion angle in the equatorial region 60°–120°, whilst 98 Σ^- decays gave 46 pions in the equatorial region.

K^- *capture in deuterium* has been discussed briefly by Case, Karplus and Yang (1956). No experimental results are yet available but deuterium bubble chambers should soon be available for this experiment.

Owing to the weak binding of the deuteron, the $(\Lambda\pi)$ and $(\Sigma\pi)$ reactions discussed above may now occur for either neutron or proton of the deuteron, for example,

$$K^- + D \begin{array}{ll} \nearrow \Lambda^0 + \pi^- + P, & \quad \ldots\ldots(4.35a) \\ \searrow \Lambda^0 + \pi^0 + N, & \quad \ldots\ldots(4.35b) \end{array}$$

and these reactions, which lead to three-particle final systems, may well be the predominant capture processes. Their observation will be of special value in providing checks on the K^-–neutron interaction which has been deduced from the analysis of K^-–proton absorption on the basis of charge independence. However, further reactions not involving a final pion are also to be expected,

$$K^- + D \begin{array}{ll} \nearrow \Sigma^- + P + 245 \text{ Mev}, & \quad \ldots\ldots(4.36a) \\ \rightarrow \Sigma^0 + N, \Sigma^0 \rightarrow \Lambda^0 + \gamma, & \quad \ldots\ldots(4.36b) \\ \searrow \Lambda^0 + N + 320 \text{ Mev}. & \quad \ldots\ldots(4.36c) \end{array}$$

A feature of special interest in deuterium K^- capture is that $T = 0$ for the deuteron, so that the final systems in these reactions necessarily have $T = \frac{1}{2}$. This implies that the reactions (*a*) and (*b*) should have relative frequency 2 : 1 for processes (4.35) and (4.36), in accord with their relative intensity in an isotopic-spin state $T = \frac{1}{2}$ (cf. eqn. (2.6)). It is of considerable importance to verify specific isotopic-spin predictions of this kind, since there is still very little quantitative evidence bearing on the assumption of charge-independence made in the Gell-Mann scheme. In the present case, unfortunately, it may be difficult to distinguish events (4.36*c*) from the Σ^0 events (4.36*b*) ; also the Λ^0 energy distributions from (4.35*b*) and (4.36*b*) overlap, all of which will confuse the verification of these isotopic-spin predictions.

However, Lee (1955) has pointed out that K^- reactions with helium would provide less ambiguous processes for study since the final system has more charged particles ; this means that the neutral secondary can often be identified by observations on charged secondaries. Several predictions for helium reactions closely related to the reactions (4.35) and (4.36) have already been discussed in § 2.5.

K^- *capture in nuclei.* First the K^-–neutron capture reactions (4.32) must be considered. These are characteristic of the $T = 1$ transitions and lead to the state

$$a_1\left[\frac{U_-(\Sigma)U_0(\pi) - U_0(\Sigma)\,U_-(\pi)}{\sqrt{2}}\right] + b_1\,U_0(\Lambda)\,U_-(\pi), \qquad \ldots\ldots(4.37)$$

neglecting reactions involving two pions. Secondly, the processes observed will differ from the K^-–nucleon capture processes as a result of the secondary interactions of the outgoing Σ and π particles with the remaining nucleons.

Since the emulsion nuclei contain roughly equal numbers of neutrons and protons the K^- particle will be incident on neutron or proton with about equal frequency. In this situation, charge independence implies the relationship (2.35),

$$n(\Sigma^+) + n(\Sigma^-) = 2n(\Sigma^0) \qquad \ldots\ldots(4.38)$$

for the Σ particles resulting from the primary process. For a nucleus containing equal numbers of neutrons and protons, the relationship is then maintained through the secondary interactions of the outgoing Σ and Λ particles. With neutrons and protons equally available, Σ particle absorption by the $\Sigma \to \Lambda$ reaction proceeds equally (in consequence of charge independence) for $\Sigma^+, \Sigma^0, \Sigma^-$ particles (similarly for any Σ production by the highly improbable $\Lambda \to \Sigma$ reaction); also charge-exchange Σ–nucleon scattering does not affect (4.38) since the $\Sigma^\pm \to \Sigma^0$ reactions then occur at the same rate as the $\Sigma^0 \to \Sigma^\pm$ reactions (this depends on charge-symmetry). This result (4.38) is a special example† of a general theorem proved by Watson (1951), and the same result will obviously hold also for the pions emitted in K⁻–nuclear absorption. The existence of this relation, a consequence of charge independence, depends on (i) the isotropy of the target nuclear system in isotopic spin space (which holds strictly only with $T = 0$ for the target nuclei), and (ii) the fact that the Σ particle has a greater isotopic spin than does the incident K^- particle. To test this relationship with emulsion data is obviously not feasible owing to the difficulty of observing Σ^0 particles; on the other hand, it provides us with a means for estimating the number of neutral $T = 1$ particles emitted in terms of the corresponding charged particles observed.

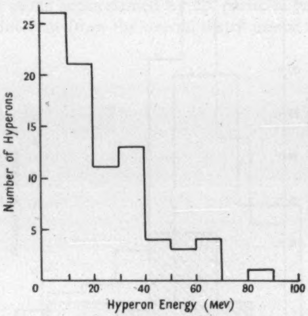


Figure 22. The energy distribution observed for 83 $\Sigma^\pm$ hyperons emitted from 617 K^- capture stars in emulsion (after Goldhaber 1956).

Figure 22 shows the energy spectrum observed for 83 charged hyperons emitted from 617 K^- capture stars. A Σ hyperon resulting from K^- capture by a stationary nucleon would receive about 10 MeV kinetic energy; the observed distribution of energies up to about 60 MeV reflects the momentum distribution of the nucleons in complex nuclei due to their Fermi motion. These charged hyperons consist about equally of Σ^+ and Σ^- particles. However, the efficiency for detecting Σ^- particles is relatively low since about 50–70% of Σ^- capture events in emulsion are believed to give rise to a zero-prong star.† Taking this loss of Σ^- particles into account, we conclude that about 150 $\Sigma^\pm$ hyperons were emitted from these 617 stars. From the relation (4.38), it may then be deduced that an additional 75 Σ^0 hyperons were also emitted. The remaining 392 events therefore gave rise to a Λ hyperon, so that the final Σ/Λ ratio for K^- capture by emulsion nuclei is about $225/392 = 0{\cdot}58$.

---

† A general proof of this relation may be given as follows. For a K^- particle incident on a given system of nucleons with random isotopic spin vectors (e.g. a nucleus of isotopic spin $T = 0$), let N_+, N_0, N_- denote the probabilities for production of $\Sigma^+, \Sigma^0, \Sigma^-$ particles, irrespective of the other products formed. The situation in which a $\overline{K}^0$ particle is incident on this random system of nucleons is charge symmetric to the situation just considered and may be obtained from it by a rotation of isotopic spin space through 180° such that the T_3 axis is reversed, which leads to the interchange of Σ^+ and Σ^- particles. From this it is clear that the probabilities for $\Sigma^+, \Sigma^0, \Sigma^-$ production by a $\overline{K}^0$ particle are simply N_-, N_0, N_+. If both K^- and $\overline{K}^0$ particles are incident with equal frequency on this nuclear system, the probabilities for $\Sigma^+, \Sigma^0, \Sigma^-$ production sum to $(N_+ + N_-), 2N_0, (N_+ + N_-)$. However, in this last mentioned situation, both initial systems are isotropic in isotopic spin and therefore the final system is necessarily isotropic in isotopic spin, which means that all Σ charge states appear with equal frequency, i.e. $N_+ + N_- = 2N_0$.

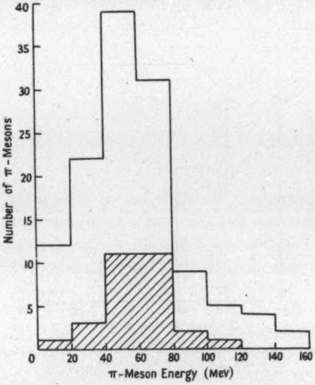


Figure 23. The energy distribution observed for 124 π mesons emitted from 410 K^- capture stars in emulsion (*after* Goldhaber 1956). The hatched area represents the π mesons emitted together with charged hyperon.

This differs considerably from the value expected for the primary K^- capture process in nuclei, following the charge-independent analysis of the data on K^- capture by protons. Averaging over K^- capture by neutron and proton, the relative probability for Σ production in the primary process is $(a_0^2+3a_1^2)/4$, whereas the probability for Λ production is $3(1{\cdot}7)b_1^2/4$. Taking the value $a_1^2/a_0^2 = 0{\cdot}6$ and the value b_1^2/a_0^2 of (4.34), this leads to an expected Σ/Λ ratio of about $5{\cdot}5(\pm 2)$. The discrepancy between these two values for this Σ/Λ ratio may be due to the fact that nuclear capture of K^- particles takes place under different conditions from capture in hydrogen—for example, if both s- and p-wave capture take place in hydrogen, then s- and p-wave capture by the nucleons in complex nuclei will generally take place with a different relative weighting from that in hydrogen, so that the charge-independence analysis could not be carried over directly to the present discussion. However, this discrepancy can also be attributed to secondary interactions of the outgoing Σ particles with the remaining nucleons of the nucleus, such as

---

† Fry, Schneps, Snow and Swami (1955 a) found Auger electrons emerging from zero-prong Σ^- capture events, with frequency about 25% of the observed Σ^- stars of one prong or more. Auger electrons are known to be emitted in about 20% of μ^- and K^- capture events, so they concluded that zero-prong Σ^- stars were perhaps a little more frequent than the observed Σ^- stars. Koshiba (1956) has also given an argument based on the observed prong distribution for Σ^- stars, from which he concluded that the zero-prong capture events were about twice as frequent as the observed stars.

$$\Sigma^- + P \rightarrow \Lambda^0 + N, \qquad \ldots\ldots(4.39)$$

which reduce the Σ/Λ ratio from the value for free nucleons.

A further estimate of this ratio can be obtained from the study of the pion energy spectrum, which is shown in figure 23 for the pions from 410 capture stars. Pions from the Λ reaction would have an initial energy of 150 Mev, compared with 90 Mev for the Σ reaction, whereas most of the pions observed have energies of 80 Mev or less. Whilst secondary interactions would give rise to a degradation of the pion energies, the pions accompanied by $\Sigma^\pm$ particles have an energy distribution not appreciably different from the overall distribution, which suggests that the Σ reaction predominates in the primary capture process. The existence of secondary interactions is directly demonstrated by the data of figure 24, which shows the total energies of associated $\Sigma+\pi$ pairs; according to this distribution, energy losses of 20–40 Mev due to excitation of the residual nucleus are typical for these outgoing particles. Taking into account the known interaction of pions with emulsion nuclei, Goldhaber (1956) estimates that the pion evidence suggests a primary Σ/Λ ratio of about 3/1 (see also Haskin, Bowen and Schein 1956), which is of the same order of magnitude as that suggested by the analysis of the $K^- + P$ data.

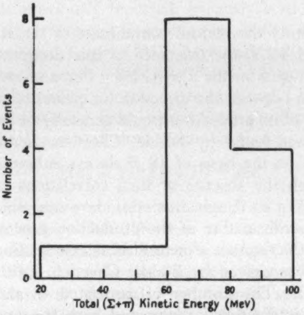


Figure 24. The distribution in energy of $(\Sigma+\pi)$ pairs emitted from K^- capture stars. The energy release for K^- capture by a stationary proton is about 104 Mev for the $(\Sigma^+ + \pi^-)$ reaction, about 95 Mev for the $(\Sigma^- + \pi^+)$ reaction.

Processes of K^- capture involving a pair of nucleons, of the type (4.36), have been neglected in the above discussion. Hyperons of much higher energy would result from these processes; reaction (4.36*a*), for example, would lead to a Σ particle of about 120 Mev. Hornbostel and Salant (1956) have described a K^- capture event involving emission of a $\Sigma^\pm$ particle of about 125 Mev; this is possible only with a capture mechanism involving two (or more) nucleons. Another event described by these authors leads to a proton of 180 Mev, which also appears to require such a capture process; several similar events have also been described by Haskin, Bowen and Schein (1956). However, from the observed hyperon and proton spectra, it seems that these processes play only a minor role in the nuclear capture of K^- particles.

§ 5. THE NEUTRAL K PARTICLES

5.1. *The θ^0 Decay*

The most frequently observed K^0 decay has been the θ^0 decay (1.2) to two charged pions with an energy release corresponding to a mass $966(\pm 10)\, m_e$ for the parent θ particle. Evidence which strongly suggests the existence of a competing decay to two neutral pions for the θ^0 particle has already been discussed in § 3.2; thus

$$\begin{tikzcd} & \pi^+ + \pi^-, & \ldots\ldots(5.1a) \\ \theta^0 \arrow[ur] \arrow[dr] & & \\ & \pi^0 + \pi^0. & \ldots\ldots(5.1b) \end{tikzcd}$$

This particle is apparently the neutral counterpart to the $K^+_{\pi 2}$ or θ^+ particle. As discussed in §§ 3.2 and 4.2.1, the existence of this decay mode (5.1*b*) implies an even parity and an even spin for the θ particles. For non-zero θ spin, the observation of some correlation between the direction (or plane) of the θ^0 decay (5.1*a*) and the direction (or plane) of its production would generally be expected. No evidence for such correlations was found by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956) on the basis of 18 θ^0 decays subsequent to the production process (3.1). Although the absence of such correlations is consistent with zero θ spin, it cannot be taken to demonstrate this, however, since lack of correlations may result from the specific nature of the production process.

An estimate for α_θ, the fraction of neutral K decays leading to the charged decay mode (5.1*a*), has also been given by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956). The number of associated θ^0 and Λ^0 decays resulting from the same $\pi^- + P$ interaction is compared with the number of $\pi^- + P$ interactions for which a Λ^0 decay was observed unaccompanied by a θ^0 decay, allowance being made for the probability of escape of the θ^0 particle from the chamber before decay. The value obtained for α_θ was as follows,†

$$\alpha_\theta = 0{\cdot}3^{+0\cdot 19}_{-0\cdot 12}. \qquad \ldots\ldots(5.2)$$

Similar values have been reported also by Blumenfeld, Booth, Lederman and Chinowsky (1956) and by Leighton (1956). The unobserved K^0 decays may be attributed both to the neutral θ^0 decay (5.1*b*) and to the existence of long-lived K^0 particles (cf. §§ 5.2 and 5.3) which decay predominantly outside the chamber. This result (5.2) gives a lower limit to the fraction of θ^0 particles with the charged decay mode (5.1*a*); the γ-ray observations (§ 3.2) require the branching ratio between (5.1*a*) and (5.1*b*) to be of general order of magnitude unity.

Since pions are Bose particles, even parity for the two final pions can occur only with an even isotopic-spin wave function; thus $T = 0$ or $T = 2$. The amplitudes with which these two isotopic-spin states contribute to the final pion state will be denoted by $A_0 \exp(i\delta_0)$ and $A_2 \exp(i\delta_2)$. According to a general theorem based on invariance under time reversal (Gell-Mann and Watson 1954, Takeda 1956), A_0 and A_2 are real quantities whilst δ_0 and δ_2 denote the pion–pion scattering phase shifts effective in the $T=0$ and $T=2$ final states. With these amplitudes and the corresponding wave functions (2.7), the branching ratio for process (5.1) may be written (Takeda 1956)

---

† *Note added in proof*: On the basis of greatly improved statistics, this group has now obtained (Rochester Conference 1957) the value $\alpha_\theta = 0{\cdot}42 \pm 0{\cdot}05$. Also photons have been observed in association with Λ^0 production events in their chamber which lead to the estimate that $7 \pm 2\%$ of the θ^0 particles produced undergo the neutral decay mode (5.1 *b*). Thus they conclude that $49 \pm 6\%$ of these θ particles undergo two-pion decay modes, the remainder undergoing decay modes with a lifetime exceeding 10^{-9} sec, in agreement with the discussion of Gell-Mann and Pais (cf. § 5.2).

$$\frac{P(\theta^0 \to \pi^0 + \pi^0)}{P(\theta^0 \to \pi^+ + \pi^-)} = \frac{A_0^2 - 2\sqrt{2}A_0 A_2 \cos(\delta_0 - \delta_2) + 2A_2^2}{2A_0^2 + 2\sqrt{2}A_0 A_2 \cos(\delta_0 - \delta_2) + A_2^2}. \qquad \ldots\ldots(5.3)$$

It is of interest to note that this branching ratio can vanish only if the phase shifts δ_0 and δ_2 happen to be equal at this pion–pion energy and if $A_0 = A_2\sqrt{2}$.

A determination of the θ^0 lifetime has been obtained by Gayther (1954) after review of the θ^0 events previously reported, by analysis of a total of 34 θ^0 decays observed in cloud chambers exposed to cosmic rays, with a result $1{\cdot}7^{+0\cdot6}_{-0\cdot4} \times 10^{-10}$ sec. Recently Snyder, Chang and Gupta (1957) have analysed a further 35 cosmic ray θ^0 decays which lead to the result $0{\cdot}8^{+0\cdot4}_{-0\cdot2} \times 10^{-10}$ sec. A similar value $0{\cdot}8^{+0\cdot3}_{-0\cdot2} \times 10^{-10}$ sec was reported by Blumenfeld, Booth, Lederman and Chinowsky (1956) on the basis of 29 θ^0 decays resulting from interactions produced by the Brookhaven π^- beam. The weighted mean of these values is then†

$$T(\theta) = 1{\cdot}0^{+0\cdot23}_{-0\cdot15} \times 10^{-10} \text{ sec.} \qquad \ldots\ldots(5.4)$$

The θ^0 lifetime is in striking contrast with the θ^+ ($K^+_{\pi 2}$) lifetime, which is longer by two orders of magnitude. This difference is in accord with the rule suggested by Gell-Mann (1955) that slow decays lead to a change $\Delta T = \frac{1}{2}$ in the total isotopic spin of the system. With $T = \frac{1}{2}$ for the θ particle, this rule allows $T=0$ or $T=1$ for the final pions. However, the odd $T=1$ isotopic spin state cannot occur for two pions with relative motion of even parity (cf. § 2.2), so that $T=0$ is required for the final pions. Although this is possible in θ^0 decay, the pions from θ^+ decay have charge $+1$ and $T_3 = +1$, and therefore cannot have $T=0$. It then appears natural for the θ^0 decay to proceed much faster than the θ^+ decay, since the latter can occur only through a violation of the $\Delta T = \frac{1}{2}$ interaction, leading to a final $T=2$ state. This interpretation has been discussed by Wentzel (1956 b), Gatto (1956 a), Kawaguchi and Nishijima (1956 a) and by Dalitz (1956 b). It should be noted that a branching ratio (5.3) of 1/2 is to be expected for a pure $T=0$ final state, although this ratio is in fact quite sensitive to a small admixture of $T=2$ state (Dalitz 1956 b).

5.2. *The Relationship between* K^0 *and* $\bar{K}^0$ *Particles*

The existence of a neutral boson, the K^0 particle, with a quite distinct antiparticle (the $\bar{K}^0$ particle) provides us with a situation different from that holding for the neutral bosons previously known—both the neutral pion and the photon are their own antiparticles. The implications of this situation for the neutral K particle were first discussed by Gell-Mann and Pais (1955). As discussed earlier

---

† *Note added in proof*: This value has been confirmed on the basis of 182 θ^0 events by the Columbia bubble chamber group (Schwartz, Rochester Conference 1957) who find a θ^0 lifetime of $0{\cdot}95 \pm 0{\cdot}08 \times 10^{-10}$ sec.

(§ 2.3), the absence of nucleon–nucleon interactions leading to two (Σ, Λ) hyperons provides the most direct empirical evidence that K^0 and $\bar{K}^0$ particles are distinct. In terms of the Gell-Mann assignments, the K^0 particle has strangeness $s = +1$, which implies[†] opposite strangeness $s = -1$ for its antiparticle; since the K^0 and $\bar{K}^0$ particles differ in the assignment of this quantum number, they must represent different states and therefore be distinct particles.

Among neutral fermions, the antineutron is known to be distinct from the neutron. This is required by the principle of conservation of baryons (§ 2.1); whereas the neutron has baryon number $B = +1$, the antineutron has $B = -1$. Since conservation of B is required to hold rigorously, there can be no interaction leading from one of these particles to the other, so that neutron and antineutron states can never be mixed. It is at this point that there is an important difference in the $K^0, \bar{K}^0$ situation; although these particles differ in value of strangeness s, conservation of s holds only for the strong and electromagnetic interactions; the weak decay interactions all lead to a change in the total strangeness s of the system. Consequently it becomes meaningful to consider mixtures of K^0 and $\bar{K}^0$ states since there do exist (weak) interactions which couple these two states; for example, since $\theta^0 \rightarrow \pi^+ + \pi^-$, the antiparticle $\bar{\theta}^0$ will similarly interact with the antiparticles of this final system, which are again $\pi^- + \pi^+$, so there will exist the chain of interactions

$$\theta^0 \longleftrightarrow \pi^+ + \pi^- \longleftrightarrow \bar{\theta}^0. \qquad \ldots\ldots(5.5)$$

The operation in which each particle of a given system is replaced by its antiparticle will be referred to as *charge conjugation* (denoted by C.C.), since this operation reverses the charge of each particle of the system. This operation necessarily also reverses the baryon number B and the strangeness s of each particle. If the initial state is described by wave function ψ let us denote the wave function of the system after this operation by ψ', thus

$$\psi \xrightarrow{\text{C.C.}} \psi'. \qquad \ldots\ldots(5.6)$$

[†] In a relativistic theory, the relation between energy E and momentum p of a system is necessarily quadratic. This means that to every solution $\psi(E) \sim e^{-iEt}$ of positive frequency, there necessarily corresponds also a solution $\psi(-|E|) \sim e^{+i|E|t}$ describing a state of negative energy.

For Fermi particles, Dirac first gave the physical interpretation for these negative energy states, by regarding the vacuum as the state in which all negative-energy levels are occupied. When a particle makes a transition from a negative-energy level to a positive-energy level, this transition is to be regarded as a pair-creation process, the creation of a particle and an antiparticle. Consequently the antiparticle is represented by the state in which all but one of the negative energy levels are occupied, i.e. by the vacuum less one negative-energy particle. From this, the fact that charge, baryon number, strangeness s and isotopic spin component T_3 are all zero for the vacuum requires that the charge, baryon number B, strangeness s and T_3 for the antiparticle should all necessarily have just opposite values from those for a particle of this species.

For Bose particles, a somewhat different interpretation leads to this same conclusion. Clearly the wave function $\phi = \psi^*(-|E|)$ is a positive frequency solution for the free particle (at least for spinless particles). For a charged particle (charge $+e$), this will still hold true in the presence of an electromagnetic field only if this state ϕ corresponds to a particle of charge $-e$: this follows from the fact that the e.m. field appears only through the combination $(\mathbf{p} - e\mathbf{A})$ and since the momentum p is reversed in the operation $\psi \rightarrow \phi$. Further, gauge invariance necessarily requires that the e.m. field (e.g. the photon) should have matrix elements connecting the particles of positive and negative charge: this means that creation of a positive and a negative particle could result from the collision of two photons. If the interactions of the theory are to allow conservation of T_3, s and B through strong and electromagnetic interactions, it therefore follows that the particle and antiparticle of a Bose field should also have opposite charge, baryon number B, strangeness s and component T_3. For an uncharged particle, it is possible that the wave function ϕ may be identical with $\psi(E)$, i.e. that the particle is its own antiparticle; however, from the above remarks, this is clearly possible only if the particle has $s = 0$, $T_3 = 0$, $B = 0$ as well as zero charge (as is the case for the π^0 meson).

Then if ψ refers to a system of total charge Q, total baryon number B and total strangeness s, then ψ' describes a system of charge $-Q$, baryon number $-B$ and strangeness $-s$. Interactions leading from state ψ to the charge conjugate state ψ' are possible only if these two states have the same charge and baryon number, which requires $Q = 0$ and $B = 0$. In this case linear combinations $\psi \pm \psi'$ of these states can be formed which are even or odd under the operation of charge conjugation, thus

$$(\psi \pm \psi') \xrightarrow{\text{C.C.}} c(\psi \pm \psi'), \qquad \ldots\ldots(5.7)$$

where $c = +1$ for the state $\psi + \psi'$ and $c = -1$ for state $\psi - \psi'$. This quantity c will be referred to as the *charge conjugation parity* of the system, and it is clear that c is only relevant to systems with $Q = B = 0$. For example, in the system $\pi^+ + \pi^-$, the operation of charge conjugation simply interchanges π^+ and π^- mesons, since these are particle and antiparticle; but this interchange has the same effect as the parity operation for their relative motion, so that the charge conjugation parity c for this system is equal to its spatial parity w.

The importance of charge conjugation parity c for these systems lies in the fact that it is natural to assume all interactions to be unchanged when the interacting particles are each replaced by their antiparticles. Indeed, it has been shown by Luders (1954) that, for any relativistically invariant theory in which the normal connection between spin and statistics (Bose statistics for integral spin particles, Fermi statistics for half-integral spins) holds, the assumption of invariance under (Wigner) time reversal (and under space reflections) involves the requirement that all interactions are invariant under this charge-conjugation operation. In this case, this property of the system with respect to particle–antiparticle interchange, the charge-conjugation parity c, will remain unchanged by any interactions which occur, whether these be strong or weak interactions; c is then a rigorous quantum number for the system.†

In order that the interaction of an electromagnetic field with charge and current should remain unchanged when particle is replaced by antiparticle, it is clear that charge-conjugation parity must be $c = -1$ for the photon, since charge and current reverse sign under charge conjugation. The neutral pion is also its own antiparticle and must have $c = +1$, since its decay leads to a final system of two photons, whose total c is $(-1)^2$.

For the neutral K particle, following (5.7), states of K^0 and $\bar{K}^0$ particles can be formed with $c = \pm 1$, thus

$$c = +1 \quad \psi(K_1{}^0) = \frac{1}{\sqrt{2}}[\psi(K^0) + \psi(\bar{K}^0)], \qquad \ldots\ldots(5.8a)$$

$$c = -1 \quad \psi(K_2{}^0) = \frac{1}{i\sqrt{2}}[\psi(K^0) - \psi(\bar{K}^0)]. \qquad \ldots\ldots(5.8b)$$

The interactions which conserve strangeness contribute the same energy to these $K_1{}^0$ and $K_2{}^0$ states since these interactions allow no cross terms between K^0 and $\bar{K}^0$

---

† *Note added in proof*: It is now known that this statement holds only for strong interactions. However, it seems still possible that product of the charge conjugation parity c and the space parity w may form a rigorous quantum number even for weak interactions. In this case the qualitative features of the physical situation would still be as discussed here, and this section would require changes only in details. The present situation for weak interactions is briefly discussed in the Appendix.

states. However, weak interactions such as (5.5) exist connecting K^0 and $\bar{K}^0$ states, and these will contribute differently to the energy of the $K_1{}^0$ and $K_2{}^0$ states ; these interactions therefore split the $K_1{}^0, K_2{}^0$ states and give rise to a mass difference δm between them. These masses will be denoted by m_1 and $m_2 = m_1 + \delta m$; this mass difference δm will be proportional to f^2, f being the interaction constant for the K^0 decay, so that δm and the lifetime width for decay will generally be of the same order of magnitude ($\sim 10^{-5}$ electron volt for the θ^0 particles).

Next, the lifetimes of the $K_1{}^0, K_2{}^0$ systems will generally be different. For the θ^0 particles, for example, it is known that the system $\pi^+ + \pi^-$ resulting from θ^0 decay has space parity $w = +1$ and therefore $c = +1$. Consequently, only the $\theta_1{}^0$ particle, which has $c = +1$, can decay to two pions,

$$\theta_1{}^0 \rightarrow \pi^+ + \pi^-. \qquad \ldots\ldots(5.9)$$

The $\theta_2{}^0$ particle can only decay to final systems with $c = -1$. Since the photon has $c = -1$, decay to $\pi^+ + \pi^- + \gamma$ would be allowed for the $\theta_2{}^0$ particle, but it seems likely that this would be an extremely slow process. Many other possibilities can be considered ; for example, the processes

$$\theta_2{}^0 \begin{cases} \rightarrow \mu^+ + \bar{\nu} + \pi^-, & \ldots\ldots(5.10a) \\ \rightarrow \mu^- + \nu + \pi^+, & \ldots\ldots(5.10b) \end{cases}$$

are conceivable $\theta_2{}^0$ decay processes, since the state

$$[\psi(\mu^- + \nu + \pi^+) - \psi(\mu^+ + \bar{\nu} + \pi^-)]$$

has $c = -1$, being clearly odd for interchange of particle and antiparticle. Although little is known concerning such interactions, comparison with the $K^+_{\mu 3}$ partial lifetime suggests that these $\theta_2{}^0$ decay processes may have lifetimes exceeding the observed $\theta_1{}^0$ decay time by a factor of at least 10^3. In this way Gell-Mann and Pais (1955) were led to predict the existence of a long-lived neutral K particle.

A similar situation will hold for the τ^0 particle, assuming it to be distinct from the θ^0 particle. From the partial lifetime observed for τ^+ decay, the $\tau_1{}^0$ decay may be expected to be slower than $\theta_1{}^0$ decay by a factor of order 10^3. Both $\tau_1{}^0$ and $\tau_2{}^0$ particles will allow a three-pion decay (with different characteristics for each case), but other decay modes may well predominate. This will be discussed briefly in § 5.3. For the remainder of this section our remarks will be confined to the θ particles.

Since strangeness conservation holds for strong interactions, the neutral θ particle state associated with the production of a Λ^0 particle, for example in the process (3.1), must necessarily have $s = +1$, and must therefore be an appropriate superposition of $\theta_1{}^0$ and $\theta_2{}^0$ states. From the expressions for the $\theta_1{}^0$ and $\theta_2{}^0$ wave functions, analogous to the equations (5.8), it follows that the wave function describing θ particles of momentum p immediately after this production process ($t = 0$) must have the following form :

$$\psi(\theta^0, t=0) = \frac{1}{\sqrt{2}}[\psi(\theta_1{}^0, 0) + i\psi(\theta_2{}^0, 0)]. \qquad \ldots\ldots(5.11)$$

After $t = 0$, the time variation for this state (for motion in free space) arises from the phase changes due to the difference in energy of the $\theta_1{}^0$ and $\theta_2{}^0$ particles and from the decay processes. Then, at time t, the wave function becomes

$$\psi(\theta^0, t) = \frac{1}{\sqrt{2}}\exp(i\omega_1 t)\,[\exp(-\tfrac{1}{2}\lambda_1 t)\,\psi(\theta_1{}^0, 0) + i\exp(it\Delta)\exp(-\tfrac{1}{2}\lambda_2 t)\psi(\theta_2{}^0, 0)], \qquad \ldots\ldots(5.12)$$

where $\omega_1 = \sqrt{(m_1{}^2 + p^2)}$, $\Delta = \omega_2 - \omega_1 \simeq m_1\,\delta m/\omega_1$ and λ_1, λ_2 are the decay constants for $\theta_1{}^0$ and $\theta_2{}^0$ decay respectively. After a flight-time long compared with 10^{-10} sec, the $\theta_1{}^0$ component of the beam will have decayed away, leaving a pure $\theta_2{}^0$ beam whose decay processes generally take place very far from the primary target. Although long-lived K^0 decays have now been observed (see § 5.3), these cannot yet be taken to demonstrate the existence of the $\theta_2{}^0$ particle, as the decays observed could still arise from long-lived τ^0 decays. A study of their nuclear interactions will be necessary to establish clearly the nature of these long-lived K^0 particles.

The nuclear interactions of the $\theta_2{}^0$ particle will be governed by strong interactions conserving s. However, the $\theta_2{}^0$ particle does not have a definite value of s but, according to (5.8*b*), consists equally of a θ^0 part with $s = +1$ and a $\bar{\theta}^0$ part with $s = -1$, thus

$$\psi(\theta_2{}^0) = [\psi(\theta^0) - \psi(\bar{\theta}^0)]/i\sqrt{2}. \qquad \ldots\ldots(5.13)$$

In a nuclear interaction of a $\theta_2{}^0$ particle, the $\bar{\theta}^0$ part can undergo an absorptive interaction resulting in Λ or Σ production, analogous to the K^- interactions (4.31), whereas the θ^0 part can have only a weak scattering or charge exchange interaction, similar to the K^+–nucleon processes (4.21). Consequently, when the $\theta_2{}^0$ beam remaining after a long flight time strikes a thin target, nuclear reactions leading to the appearance of Λ and Σ particles are to be expected. These nuclear reactions are not possible for the θ particle state (5.11) arising initially from the production process since this is a pure $s = +1$ state containing no component with negative strangeness $s = -1$, so that observation of these secondary hyperons at distances far (in terms of their lifetimes) from the primary target would provide a striking demonstration of the character of the long-lived $\theta_2{}^0$ particles. Further, if the target is thick enough to remove an appreciable part of the $\bar{\theta}^0$ component of (5.13) (the mean free path is about 100 g cm^{-2} according to the K^- absorption data discussed in § 4.7), only the θ^0 component of the $\theta_2{}^0$ beam, being relatively weakly affected, will pass on beyond the target with appreciable amplitude. Since this θ^0 beam consists equally of a $\theta_1{}^0$ and a $\theta_2{}^0$ part (cf. equation (5.11)), $\theta_1{}^0$ decay events will now appear again immediately behind this secondary target. A preliminary discussion of these effects has recently been given by Pais and Piccioni (1955), and a number of groups have experiments of this general type in progress at present (cf. § 5.3.3).

The passage of a $\theta_2{}^0$ particle beam through absorbing material has been discussed in more detail by Case (1956). The $\bar{\theta}^0$ component of the incident $\theta_2{}^0$ beam is removed by nuclear absorption as the beam moves through the material, leaving the weakly interacting θ^0 component; however, since $\theta_1{}^0$ decay occurs rapidly relative to the rate of $\bar{\theta}^0$ absorption, this θ^0 component rapidly loses its $\theta_1{}^0$ part through decay, so that the $\theta_2{}^0$ beam is continually being partially regenerated. To investigate the manner in which these components vary through the material, we denote the wave function at time t by

$$[a(t)\,\psi(\theta^0)+\bar{a}(t)\,\psi(\bar{\theta}^0)]\exp(i\omega_1 t)\equiv[a_1(t)\,\psi(\theta_1{}^0)+a_2(t)\,\psi(\theta_2{}^0)]\exp(i\omega_1 t) \qquad \ldots\ldots(5.14)$$

where, from equation (5.8),

$$a_1(t)=[a(t)+\bar{a}(t)]/\sqrt{2},\quad a_2(t)=[a(t)-\bar{a}(t)]/i\sqrt{2}. \qquad \ldots\ldots(5.15)$$

Now the time variation of these coefficients will result from the following causes:

(*a*) Difference between the phase of $a_1(t)$ and $a_2(t)$ due to the $\theta_1{}^0, \theta_2{}^0$ mass difference δm. This contributes

$$\left(\frac{da_1(t)}{dt}\right)_{\text{phase}}=0,\quad \left(\frac{da_2(t)}{dt}\right)_{\text{phase}}=i\Delta a_2(t), \qquad \ldots\ldots(5.16a)$$

where $\Delta = m_1\,\delta m/\omega_1$ as before.

(*b*) The nuclear absorption of the $\bar{\theta}$ component and the scattering-out of the θ component. Since the θ scattering is predominantly elastic (cf. § 4.5), the scattering-out loss should be included only for conditions of good geometry; with poor geometry the scattered θ particles will still contribute since they remain θ particles after scattering. This contributes

$$\left(\frac{da(t)}{dt}\right)_{\text{sc}}=-\frac{1}{2\tau_s}\,a(t),\quad \left(\frac{d\bar{a}(t)}{dt}\right)_{\text{abs}}=-\frac{1}{2\tau_a}\,\bar{a}(t), \qquad \ldots\ldots(5.16b)$$

where τ_s denotes the mean time-of-flight for the θ^0 particle before scattering, and τ_a denotes the mean time-of-flight for the $\bar{\theta}^0$ particle before nuclear absorption.

(*c*) The decay of the $\theta_1{}^0$ component,

$$\left(\frac{da_1(t)}{dt}\right)_{\text{decay}}=-\frac{1}{2\tau}\,a_1(t),\quad \left(\frac{da_2(t)}{dt}\right)_{\text{decay}}=0, \qquad \ldots\ldots(5.16c)$$

$\tau = 1/\lambda_1$ being the mean lifetime for $\theta_1{}^0$ decay.

Collecting together these contributions (5.16) and taking account of the relation (5.15) leads us to the following differential equations for the amplitudes $a_1(t)$ and $a_2(t)$

$$\begin{aligned}\frac{da_1(t)}{dt}&=\left(-\frac{1}{2\tau}-\frac{1}{4\tau_a}-\frac{1}{4\tau_s}\right)a_1(t)+i\left(\frac{1}{4\tau_a}-\frac{1}{4\tau_s}\right)a_2(t),\\ &=\alpha a_1(t)+i\lambda a_2(t). \qquad \ldots\ldots(5.17a)\\ \frac{da_2(t)}{dt}&=-i\left(\frac{1}{4\tau_a}-\frac{1}{4\tau_s}\right)a_1(t)+\left(i\Delta-\frac{1}{4\tau_a}-\frac{1}{4\tau_s}\right)a_2(t),\\ &=-i\lambda a_1(t)+\beta a_2(t). \qquad \ldots\ldots(5.17b)\end{aligned}$$

The solution of these equations (5.17) with values $a_1(0), a_2(0)$ corresponding to the initial beam then provides the amplitudes $a_1(t), a_2(t)$ describing the composition of the beam at any point during its passage through the material.

For example, consider the case of an absorber of thickness $d = vT$ (v, the θ particle's velocity) placed in the path of a pure $\theta_2{}^0$ beam. When the beam reaches the absorber ($t = 0$), $a_1(0) = 0$, $a_2(0) = A$, and the appropriate solution of (5.17) is

$$\begin{aligned}a_1(t)&=-i\lambda A\,[\exp(-\mu_2 t)-\exp(-\mu_1 t)]/(\mu_2-\mu_1),\\ a_2(t)&=A[(\alpha+\mu_2)\exp(-\mu_2 t)-(\alpha+\mu_1)\exp(-\mu_1 t)]/(\mu_2-\mu_1), \qquad \ldots\ldots(5.18)\end{aligned}$$

where $\mu_2, \mu_1 = [-(\alpha+\beta) \pm \sqrt{\{4\lambda^2+(\alpha-\beta)^2\}}]/2$. The absorption curve for the $\theta_2{}^0$ beam is then given by

$$\frac{I(d)}{I(0)} = \left|\frac{a_2(d/v)}{a_2(0)}\right|^2, \qquad \ldots\ldots(5.19)$$

the initial intensity of $\theta_1{}^0$ decays behind the target being proportional to $|a_1(d/v)|^2$. Experimental measurements (cf. § 5.3) relating to this absorption curve have recently been carried out by Panofsky and Fitch (1956, unpublished). Finally, the distribution in the production of (for example) Λ^0 particles by this beam within the absorber is proportional to $|\bar{a}(t)|^2$, where $\bar{a}(t)$ may be obtained from (5.15) and (5.18).

Case (1956) has also discussed the case of θ^0 particles created within absorbing material (e.g. within a xenon bubble chamber). Serber (see Pais and Piccioni 1955) has pointed out that if the mass difference δm is sufficiently large relative to the $\theta_1{}^0$ lifetime width, nuclear absorptions leading to hyperons can then occur within the material rather close to the primary target, after flight times of order $2\pi/\Delta \lesssim 10^{-10}$ sec. To show this, we consider the extreme case of a very large mass difference, $\Delta \gg 1/\tau = \lambda_1$, so that this effect can occur for flight times so short that the effect of the decay and absorption processes in (5.17) may be neglected. In this case equation (5.12) applies and, with neglect of decay, this reduces to

$$\frac{1}{\sqrt{2}}[\psi(\theta_1{}^0)+i\exp(it\Delta)\,\psi(\theta_2{}^0)]\exp(i\omega_1 t)$$
$$= \left[\frac{1+\exp(it\Delta)}{2}\psi(\theta^0)+\frac{1-\exp(it\Delta)}{2}\psi(\bar{\theta}^0)\right]\exp(i\omega_1 t). \qquad \ldots\ldots(5.20)$$

At $t=0$, this represents $\psi(\theta^0)$ corresponding to the production of a state with positive strangeness $s=+1$; however, at time $t=\pi/\Delta$, $e^{it\Delta}=-1$ and the function (5.20) has become $\psi(\bar{\theta}^0)$. Clearly this wave function oscillates between $\psi(\theta^0)$ and $\psi(\bar{\theta}^0)$, at least until $\theta_1{}^0$ decay begins to be important, so that the $\bar{\theta}^0$ state is effective half the time on average and can initiate nuclear reactions leading to the appearance of hyperons. If the oscillatory variation of these reactions with distance could be measured, this would allow an estimate of δm despite its extreme smallness (Pais and Piccioni 1955, Case 1956).

5.3. *Other* K^0 *Particles and* K^0 *Decay Modes*

It has been found by various cloud chamber groups (cf. Thompson 1956) that about 5–10% of the V^0 decays observed in cosmic-ray experiments correspond to $Q(\pi,\pi)$ values (Q values obtained assuming the event to represent a two-body decay to $\pi^+ + \pi^-$) quite incompatible with the value 215(±5) MeV characteristic of the $\theta_1{}^0$ decay (5.9). For these 'anomalous V^0 decays', it has not generally been possible to identify the charged secondary particles; in several favourable cases momentum and ionization measurements on a secondary track have enabled it to be distinguished as an electron (see below), whilst the observation of several secondary π–μ decays requires some secondaries to be pions. Since Gell-Mann and Pais have predicted the existence of a long-lived $\theta_2{}^0$ particle with decay modes necessarily distinct from the $\theta_1{}^0$ decay, it is natural to attribute some of these 'anomalous V^0 decay' events to $\theta_2{}^0$ decay. In fact, owing to the close relationship of K^0 and K^+ particles as an isotopic-spin doublet in the Gell-Mann scheme, we may expect that both the $\theta_1{}^0$ and $\theta_2{}^0$ particles should have decay modes closely related to the minor decay modes of the K^+ particles ; for example, the following neutral K decays may be expected to occur, for either the τ^0 or θ^0 particles (or both),

$$K_1{}^0, K_2{}^0 \begin{cases} \rightarrow \mu^- + \nu + \pi^+, \\ \rightarrow \mu^+ + \bar{\nu} + \pi^-, \end{cases} \qquad \ldots\ldots(5.21)$$

$$K_1{}^0, K_2{}^0 \begin{cases} \rightarrow e^- + \nu + \pi^+, \\ \rightarrow e^+ + \bar{\nu} + \pi^-. \end{cases} \qquad \ldots\ldots(5.22)$$

Similarly, whether or not the τ particle is distinct from the θ particle, the occurrence of decay modes involving three pions may also be anticipated,

$$\tau^0 \begin{cases} \rightarrow \pi^+ + \pi^- + \pi^0, & \ldots\ldots(5.23a) \\ \rightarrow \pi^0 + \pi^0 + \pi^0. & \ldots\ldots(5.23b) \end{cases}$$

The present evidence on 'anomalous V^0 decays' from cosmic-ray and accelerator experiments will now be described and discussed in terms of the characteristics of these decay modes (5.21), (5.22) and (5.23).

5.3.1. *The τ^0 particle.*

The appearance of τ^0 decay (5.23*a*) will be quite similar to θ^0 decay. However, the apparent Q value $Q(\pi^+, \pi^-)$, which denotes the total kinetic energy of the π^+ and π^- mesons in their centre-of-mass system, will no longer be unique since it is related to the kinetic energy t_0 carried away by the neutral pion in the τ^0 rest system by the relation

$$Q(\pi^+, \pi^-) = \sqrt{[(M_\tau - m_{\pi^0})^2 - 2M_\tau t_0]} - 2m_\pi. \qquad \ldots\ldots(5.24)$$

Values of $Q(\pi^+, \pi^-)$ between 0 and 80 MeV may occur, with relative frequencies depending on the distribution of the π^0 kinetic energy t_0 (which should be quite similar to the π^- energy distribution in τ^+ decay).

Some of the anomalous V^0 decays whose secondary tracks could be pions give $Q(\pi, \pi)$ values lying in this range and could represent τ^0 decay events. However, many others correspond to $Q(\pi, \pi)$ values lying in the range 100–200 MeV, so that only a relatively small proportion of 'anomalous V^0 events' could be due to τ^0 decay.

A V^0 decay event recently observed by the C.E.R.N. group (Cooper *et al.* 1956) does appear to require the existence of τ^0 decay for its interpretation. This event involves four secondary tracks, two of which are identified as belonging to an electron pair ; the other two tracks are short and only lower limits can be given for their momenta. The nuclear interaction from which the V^0 particle emerges is seen, so that the V^0 line-of-flight is known ; further, the momentum of the incident particle causing this interaction is measured to be about 1 GeV/*c*.

The simplest interpretation of this event is as τ^0 decay (5.23*a*), followed by the π^0 decay, $\pi^0 \rightarrow \gamma + e^+ + e^-$. The observed directions and momenta are compatible with this interpretation provided the τ^0 momentum lies in the range 600–1400 MeV/*c*, and in fact τ^0 momenta of up to about 900 MeV/*c* are energetically possible from the observed nuclear interaction with a primary particle of the measured momentum.

In terms of known particles this V^0 event could alternatively be interpreted as a very rare mode of θ decay,

$$\theta^0 \rightarrow \pi^+ + \pi^- + e^+ + e^-, \qquad \ldots\ldots(5.25)$$

a radiative θ decay in which the photon is internally converted. However, this interpretation requires a momentum of 1400 MeV/c for the θ particle, which is considered incompatible with the observed nuclear interaction. Many other conceivable interpretations (e.g. $\Lambda^0 \rightarrow P + \pi^- + e^+ + e^-$) are shown to be excluded, so that this event may be considered to provide strong evidence for the existence of a τ^0 decay.

For the τ^0 particle, the considerations of Gell-Mann and Pais (§ 5.2) will again apply (Snow 1956, Dalitz 1956 b) and the 'neutral τ particles' with distinct mass and lifetime will be the $\tau_1{}^0$ and $\tau_2{}^0$ particles which have charge-conjugation parity $c = +1$ and -1 respectively. For each of these particles, a decay to three pions will be permitted. If we now denote by l the angular momentum of the π^+ and π^- in their centre-of-mass system and by L the angular momentum of the π^0 meson in the τ^0 rest system, then $l = L$ is necessary for a τ^0 particle of zero total angular momentum (which necessarily has odd parity (cf. § 4.3)). Under the operation of charge conjugation, π^+ and π^- mesons are interchanged, the π^0 meson being unaffected; however, this interchange is identical with the parity operation in the relative motion of the π^+ and π^- mesons. From this it follows that the charge conjugation parity c of this state is $(-1)^l$. Snow has pointed out that this interchange of the π^+ and π^- mesons can also be achieved by a suitably chosen 180° rotation of the pion isotopic spins (however this rotation also reverses the π^0 wave function). For a system which has isotopic spin T and component $T_3 = 0$ (as is the case here), this rotation reproduces the same state with a factor $(-1)^T$, since the system is described by a spherical harmonic $Y_T{}^0$ in isotopic-spin space; however, the preceding remark shows that this factor is also $-c$, so that

$$c = -(-1)^T = (-1)^l, \qquad \ldots\ldots(5.26)$$

where T denotes the total isotopic spin of the final pions.

For the $\tau_1{}^0$ particle with $c = +1$, therefore, the angular momenta l, L can only take even values $l = L = 0, 2, \ldots$, which are just the configurations allowed for the τ^+ decay; similarly only total isotopic spin $T = 1$ or 3 is allowed for the final pions. If the $\Delta T = \frac{1}{2}$ rule (cf. §§ 4.4 and 5.1) holds for this decay, then only $T = 1$ is permitted for the final pions, which would imply branching ratio 2 : 3 between modes (5.23*a*) and (5.23*b*); this would also require the partial lifetime for 3π decay to be the same ($\sim 10^{-7}$ sec) for the $\tau_1{}^0$ particle as for the τ^+ particle.

On the other hand, for the $\tau_2{}^0$ particle ($c = -1$), only odd values $l = L = 1, 3, \ldots$ are allowed so that the outgoing pions face centrifugal barriers owing to their angular momentum, even in the simplest configuration. This means that the partial lifetime for $\tau_2{}^0$ decay to pions will be correspondingly lengthened (by at least $\sim 10^2$) and that other decay modes (such as $e^{\pm} + \nu + \pi^{\mp}$, for example) will probably be dominant.

Even for the $\tau_1{}^0$ particle, other decay modes may well compete, or predominate. For example, Schwinger (1956) has suggested that the K_{e3}^+ decay mode may belong to the τ^+ particle (cf. § 7.4); in this case it would be reasonable to expect the two decay modes

$$\tau_1^0 \begin{cases} \to e^- + \nu + \pi^+, \\ \to e^+ + \bar{\nu} + \pi^-, \end{cases} \qquad \ldots\ldots(5.27)$$

to occur with equal frequency at a rate at least comparable with the 3π decay. It should be added that, in this case, the τ_2^0 particle would also be expected to have the decay modes (5.27) with comparable partial lifetime.

5.3.2. K⁰ *decays with identified secondaries.*

The earliest ' anomalous V⁰ decay ' which allowed identification of the secondary particles was an event described by Cowan (1954). The positive secondary of the event itself underwent a decay consistent with the interpretation of π–μ decay and was therefore identified as a π^+ meson. From the momentum and ionization of the negative secondary, interpretation as a negative proton or as an electron was indicated. With the former possibility, the value $Q(\bar{P}, \pi^+) = 12(\pm 4)$ MEV was obtained, so that the event could not be interpreted as the decay of an anti-Λ^0 particle, except perhaps as some rare mode involving an additional photon, for example. The interpretation now preferred is that the negative secondary was an electron; the Q value $Q(e^-, \pi^+)$ is then $225(\pm 20)$ MEV. In terms of known particles, this decay must be at least a three-body decay and the simplest interpretation is that it represents a neutral K decay

$$K^0 \to e^- + \pi^+ + \nu \ (?). \qquad \ldots\ldots(5.28)$$

Since the known K mesons have integral spin, it is natural to identify the neutral particle in this decay (5.28) with the neutrino.

Further V⁰ decays with electron secondaries have been observed by Block, Harth, Blevins and Slaughter (1956) in a cloud chamber exposed to a 1·9 GEV π^- beam at Brookhaven. From ionization and momentum the positive secondary of one V⁰ event was identified as a positron, the negative secondary having a mass less than $500m_e$; the large opening angle of 68° makes it very improbable that this represents a π^0 decay ($\pi^0 \to \gamma + e^+ + e^-$) although the possibility cannot be excluded (Peyrou 1956). With the interpretation

$$K^0 \to e^+ + \pi^- + \bar{\nu}, \qquad \ldots\ldots(5.29)$$

this event gives a value $Q(e^+\pi^-) = 48(\pm 5)$ MEV. A second event gave an electron, identified from ionization and momentum measurements. For this event the positive secondary had high momentum (~ 1200 MEV/c) and could not be identified; if this is assumed to be an example of the process (5.28), a value $Q(e^-, \pi^+) = 18 \pm 4$ MEV is obtained.

A clear example of the decay mode (5.28) has recently been reported by d'Andlau, Armenteros, Astier, de Staebler, Gregory, Leprince-Ringuet, Muller, Peyrou and Tinlot (1956). From the occurrence of π–μ decay, the positive secondary is identified as π^+ meson; the negative secondary had momentum $395(\pm 1{\cdot}5)$ MEV/c and an ionization consistent only with its interpretation as an electron track. For this event $Q(e^-, \pi^+)$ had the value $92(\pm 9)$ MEV. This Q value differs from that obtained by Cowan and from that of Block *et al.*, which indicates again that these events represent decay into at least three particles. In terms of known particles, their interpretation as K^0 decay is the only three-body possibility and this is satisfactory in view of the close relationship between these decay modes (5.28) and (5.29), and the $K^{\pm}_{e3}$ decay mode.

In a recent experiment designed to detect the decay of long-lived K^0 particles, Lande, Booth, Impeduglia, Ledermann and Chinowsky (1956) have obtained a number of 'anomalous V^0 events' which are inconsistent with the τ^0 decay (5.23*a*) and appear to represent predominantly the decay modes (5.22). This experiment will be described in the following section.

5.3.3. *Long-lived* K^0 *decays*.

Following the prediction of a long-lived $\theta_2{}^0$ particle by Gell-Mann and Pais (cf. § 5.2), a Columbia group (Lande, Booth, Impeduglia, Ledermann and Chinowsky 1956) exposed a 36 in. magnet cloud chamber at a distance of 6 metres (about 10^{-8} sec flight time) to the neutral radiation emitted from a copper target at an angle of 68° to an incident 3 GeV external proton beam. A lead collimator four feet thick is followed by a magnet system to sweep charged particles out of the beam; a lead filter (1·5 in. thick) was placed before the collimator to reduce the photon intensity reaching the chamber from the machine. About 10 000 neutrons passed through the chamber each pulse, but, owing to their large angle relative to the incident protons, their energies were mostly below the pion production threshold.

Measurements on 23 V^0 events observed in the chamber have been described in the preliminary report. None of these events is compatible with normal $\theta_1{}^0$ decay, and only one event is coplanar with the line of flight, indicating that three-body decay modes are the rule. Three of the negative secondaries were identified as electrons, with high probability; two other negative secondaries were definitely not electrons but represented either π^- or μ^- mesons. Three positive secondaries were considered to be most probably π^+ mesons; none of the charged secondaries was compatible with a particle of nucleonic mass. For 12 of these 23 events, interpretation as τ^0 decay was explicitly excluded, either because of an identified electron secondary or because the value of $Q(\pi, \pi)$ obtained lay outside the allowed range of 0–80 MeV.

For an assumed incident K^0 velocity, a primary mass value could be calculated for each event for any assumed three-body decay mode. Lande *et al.* have plotted the average mass value obtained for 15 events as function of the velocity of the parent particle, for each of the decay modes $(e^- + \pi^+ + \nu)$ and $(\mu^- + \pi^+ + \nu)$; other decay modes such as $(\mu^+ + \pi^- + \bar{\nu})$ or $(\mu^+ + \pi^0 + e^-)$ give curves very close to the latter curve. Taken together, these data lead to a reasonable K^0 mass value (see figure 25) only for the $(e^- + \pi^+ + \nu)$ scheme and it is concluded that this is the predominant decay mode present. This assumption also gives reasonable values for the mean velocity of the K^0 particles; Sternheimer has calculated the K meson energy spectrum at 68° with various assumptions concerning the collision mechanism, each of these cases leading to a spectrum peaking at about 100 MeV. This conclusion does not exclude the possibility that $(\mu\nu\pi)$ and $(\pi^+\pi^-\pi^0)$ decays might amount to as much as 30% of these anomalous V^0 events.

By comparison of the observed rate of occurrence of these V^0 events with what little is known concerning the production cross section for K^0 mesons under the present conditions, upper and lower limits may be placed on the lifetime T of these particles, namely 10^{-6} sec $\geqslant T \geqslant 3 \times 10^{-9}$ sec. A weaker lower limit of $T > 10^{-9}$ sec may be deduced from the fact that the decay events are uniformly distributed in the chamber.

This experiment has established the existence of a neutral K particle with long lifetime and with modes of decay distinct from $\theta_1{}^0$ decay. Despite this qualitative accord with the Gell-Mann–Pais prediction of a long-lived $\theta_2{}^0$ particle, the nature of this K particle is not made clear in this experiment. As Lande *et al.* emphasize, the observation of long lifetime and of 'anomalous decay modes' could also result if the neutral beam at the position of the cloud chamber consisted mainly of neutral τ particles with a dominant decay mode $(e^- + \nu + \pi^+)$.

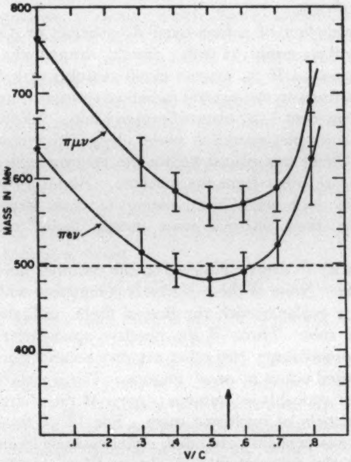


Figure 25. Average calculated primary mass plotted against velocity of primary particle for assumed decay schemes $(\pi^+ e^- \nu)$ and $(\pi^+ \mu^- \nu)$. The arrow denotes the peak of the velocity spectrum calculated for K^0 particles by Sternheimer from phase space considerations for the production process. The vertical bars are average deviations of the mean (after Lande, Booth, Impeduglia, Ledermann and Chinowsky 1956).

However, there are now several experimental results which suggest that the long-lived neutral K particles do have a component of negative strangeness. Fry, Schneps and Swami (1956 b), in examining emulsions exposed to the Berkeley K^- beam, found four neutral-induced stars from each of which emerged a particle of negative strangeness, a Λ hyperfragment, a hyperon or a K^- particle. These authors argued that these events were too many to be due to $\bar{K}^0$ particles produced by charge exchange from the K^- beam (also, one star was too energetic for this explanation), or to energetic neutrons coming from the internal target, and pointed out that they could be accounted for very readily as due to nuclear interactions induced by long-lived K_2^0 particles† coming directly from the internal target (flight time about 10^{-8} sec). Also Lander, Fowler and Powell (1956) have recently exposed a propane bubble chamber to a neutral K beam at Berkeley and have observed a number of Λ^0 decays occurring in the chamber despite its distance (3 metres) from the external target (Al bombarded with negative pions). Since all charged particles are swept out of the final beam by a magnetic field, these Λ^0 decays have also been attributed to the interaction of the neutral K beam with the material of the bubble chamber. If this preliminary interpretation of these observations is confirmed, they will provide a convincing demonstration that these long-lived neutral K particles contain a component of negative strangeness, so that these particles will have all the essential properties of the θ_2^0 particle predicted by Gell-Mann and Pais.

5.3.4. *The time dependence of 'anomalous' θ^0 and τ^0 decay*

As a further consequence of symmetry under particle–antiparticle conjugation Treiman and Sachs (1956) have pointed out that in general these 'anomalous' decays will have a time dependence more complicated than the usual exponential law. The observation of such effects would be important as a direct indication of the K_1^0, K_2^0 situation, and would also allow a sensitive estimate of the mass difference between K_1^0 and K_2^0 particles.

For definiteness, we shall consider the $(\pi e \nu)$ decay mode of a neutral K particle. Then, assuming the charge-conjugation parity c to be a rigorous quantum number of the system ($c = +1$ for the K_1^0 particle, $c = -1$ for the K_2^0 particle), the final state reached in K_1^0 or K_2^0 decay must have the same value of c as the parent particle. Thus for the K_1^0 particle this decay leads to the symmetrical final state,

$$\psi(K_1^0) \to a_1[\psi(e^+ + \bar{\nu} + \pi^-) + \psi(e^- + \nu + \pi^+)], \qquad \ldots\ldots(5.30a)$$

where a_1 denotes the decay amplitude for this process, whilst the K_2^0 decay leads to the antisymmetrical final state

$$\psi(K_2^0) \to -ia_2[\psi(e^+ + \bar{\nu} + \pi^-) - \psi(e^- + \nu + \pi^+)]. \qquad \ldots\ldots(5.30b)$$

These amplitudes a_1 and a_2 are in fact restricted to real values in consequence of the natural requirement that these interactions be invariant under (Wigner) time reversal, as pointed out by Treiman and Sachs. It may be noted that, if a_1 and a_2 are equal, the K^0 state (5.11) has no coupling with the $(e^- + \nu + \pi^+)$ system.

The time-variation of an initial K^0 state produced in association with a hyperon has already been given in equation (5.12),

$$\frac{1}{\sqrt{2}}\{\psi(K_1^0)\exp(-\lambda_1 t/2) + i\psi(K_2^0)\exp(it\Delta)\exp(-\lambda_2 t/2)\}\exp(i\omega_1 t). \qquad \ldots\ldots(5.31)$$

---

† *Note added in proof*: Lande, Lederman and Chinowsky (1957) have also observed in the helium of their cloud chamber exposed to the Brookhaven neutral K beam a neutral-induced star from which a Σ^- particle emerged. Also Baldo-Ceolin, Dilworth, Fry, Greening, Huzita, Limentani and Sichirollo (1957) and Ammar, Friedman, Levi-Setti and Telegdi (1957) have described a large number of similar neutral-induced stars found in emulsions exposed to a similar neutral K beam at Berkeley. In each of these stars just one strange particle emerged; most frequently the emergent strange particle was of negative strangeness, Λ hyperfragments, Σ particles and K^- particles being identified, but several of these stars gave a particle of positive strangeness, a K^+ particle.

Treiman and Sachs then show that the rates at which $(e^- + \nu + \pi^+)$ and $(e^+ + \bar{\nu} + \pi^-)$ decays appear at time t are proportional to the squares of the amplitudes of these two states in (5.31), after substituting the expressions (5.30), thus

$$\begin{aligned} R(e^+ + \bar{\nu} + \pi^-) &= \tfrac{1}{2}\,|a_1 \exp(-\lambda_1 t/2) + a_2 \exp(it\Delta)\exp(-\lambda_2 t/2)|^2 \\ &= \tfrac{1}{2}\{a_1^2 \exp(-\lambda_1 t) + a_2^2 \exp(-\lambda_2 t) \\ &\qquad + 2a_1 a_2 \cos(t\Delta)\exp[-(\lambda_1+\lambda_2)t/2]\}, \quad \ldots\ldots(5.32a) \end{aligned}$$

$$\begin{aligned} R(e^- + \nu + \pi^+) &= \tfrac{1}{2}\,|a_1 \exp(-\lambda_1 t/2) - a_2 \exp(it\Delta)\exp(-\lambda_2 t/2)|^2 \\ &= \tfrac{1}{2}\{a_1^2 \exp(-\lambda_1 t) + a_2^2 \exp(-\lambda_2 t) \\ &\qquad - 2a_1 a_2 \cos(t\Delta)\exp[-(\lambda_1+\lambda_2)t/2]\}. \quad \ldots\ldots(5.32b) \end{aligned}$$

In addition to the exponential terms describing the decay of K_1^0 and K_2^0 states, these expressions also contain an interference term with an exponential factor diminishing at the average rate $(\lambda_1 + \lambda_2)/2$ and an oscillatory factor depending on the energy difference between the K_1^0 and K_2^0 particles. Although the total rate of decay to $(\pi e \nu)$ final states,

$$R(e^+ + \bar{\nu} + \pi^-) + R(e^- + \nu + \pi^+) = a_1^2 \exp(-\lambda_1 t) + a_2^2 \exp(-\lambda_2 t) \quad \ldots\ldots(5.33)$$

follows the simple exponential law appropriate to the mixed K_1^0 and K_2^0 decay, the ratio $R(e^+ + \bar{\nu} + \pi^-)/R(e^- + \nu + \pi^+)$ varies with time as the decay proceeds.

The special case where $a_1 = a_2$ is of particular interest. With this relation the K^0 state (5.11) is coupled only with the $(e^+ + \bar{\nu} + \pi^-)$ system, the $\bar{K}^0$ state only with the $(e^- + \nu + \pi^+)$ system. This situation corresponds to the assumption generally made in the discussion of K decay modes involving neutrinos (Sachs 1955, Schwinger 1956) that the K^+ and K^0 decay couplings differ only in the charge of the pion involved, thus

$$K^+ \rightarrow e^+ + \bar{\nu} + \pi^0, \quad \ldots\ldots(5.34a)$$

$$K^0 \rightarrow e^+ + \bar{\nu} + \pi^-, \quad \ldots\ldots(5.34b)$$

with a similar relationship between K^- and $\bar{K}^0$ decay couplings. With $a_1 = a_2$, the rate $R(e^- + \nu + \pi^+)$ must necessarily be zero at the origin $(t = 0)$, where a pure K^0 state $(s = +1)$ is produced in association with a hyperon (or by charge exchange from a K^+ particle). As time proceeds, the K_1^0 component of this state decays out and the appearance of $(e^- + \nu + \pi^+)$ decays measures the strength of the $\bar{K}^0$ component which then arises in the system. The time dependence of the electron and positron decay modes in the case $a_1 = a_2$ is shown in figure 26 for two values of the energy difference $\Delta = m_1\,\delta m/\omega_1$. If these decay modes belong to the θ particle, this effect should be observable without great difficulty even for mass differences as large as ten times the θ_1^0 lifetime width (which is $\sim 4 \times 10^{-6}$ ev) or as small as one-tenth of this width. If these decay modes belong to the τ particle, there are still more uncertainties in the situation since the τ_1^0 and τ_2^0 lifetimes may be comparable; however, it is clear that it will still be of very great interest to determine the ratio of electron to positron decays as function of the flight time of the neutral beam.

Exactly similar effects are to be expected for K^0 decays leading to the systems $(\mu^+ + \bar{\nu} + \pi^-)$ and $(\mu^- + \nu + \pi^+)$. A time dependence of the e^+/e^- or μ^+/μ^- ratio in the decay of the neutral K beam can be absent only if these ratios are observed to be unity, which corresponds to $a_1 = 0$, or $a_2 = 0$, or to a mass difference Δ so large that the oscillatory terms give an unobservable fine structure, only average values being measurable.

§ 6. THE DECAY AND NUCLEAR INTERACTIONS OF HYPERONS

6.1. *The Decay Modes of Λ and Σ Hyperons*

Decay of the Λ^0 particle. From the observed distribution of $\Lambda^0 \to P + \pi^-$ events in cloud chambers, many groups have used the statistical procedures of Bartlett to estimate the mean lifetime for Λ^0 decay. Page (Page and Newth 1954, Page 1954) has summarized the cosmic-ray data up to 1955, giving the lifetime

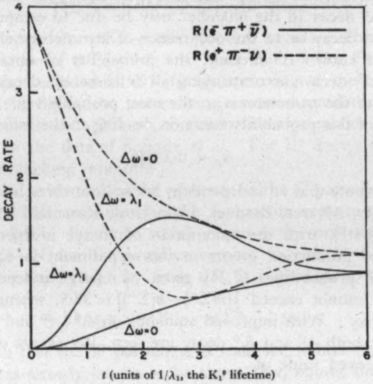


Figure 26. The decay rate of the K^0 particle into the modes $(e^+\bar{\nu}\pi^-)$ and $(e^-\nu\pi^+)$ as function of time under the assumption $a_1 = a_2$. It is assumed that $\lambda_2 \ll \lambda_1$. The two curves correspond to different choices for the $K_1{}^0$, $K_2{}^0$ energy difference Δ (*after* Treiman and Sachs 1956).

$$T(\Lambda^0) = 3{\cdot}7^{+0{\cdot}6}_{-0{\cdot}5} \times 10^{-10} \text{ sec} \qquad \ldots\ldots(6.1)$$

on the basis of 150 Λ^0 events. Blumenfeld, Booth, Lederman and Chinowsky (1956) have recently reported a somewhat shorter lifetime,† not inconsistent with (6.1),

$$T(\Lambda^0) = 2{\cdot}8^{+0{\cdot}5}_{-0{\cdot}4} \times 10^{-10} \text{ sec} \qquad \ldots\ldots(6.2)$$

on the basis of 65 Λ^0 decays observed in a 36-in. multiplate cloud chamber exposed to a Brookhaven π^- beam.

In view of the well-known charge-exchange reaction $\pi^- + P \to \pi^0 + N$, it was natural to expect a competing neutral decay-mode for the Λ^0 particle, thus

$$\Lambda^0 \begin{cases} \to N + \pi^0, & \ldots\ldots(6.3a) \\ \to P + \pi^-. & \ldots\ldots(6.3b) \end{cases}$$

---

† *Note added in proof*: This result (6.2) has been confirmed by the Columbia bubble chamber group (Schwartz, Rochester Conference 1957) who give the result $2{\cdot}77 \pm 0{\cdot}20 \times 10^{-10}$ sec on the basis of $304\Lambda^0$ events.

The existence of this neutral decay mode provides the most natural explanation for the observations on γ rays originating downstream from a target bombarded by high-energy protons, which have already been described in § 3.2.

An estimate of α_Λ, the fraction of Λ^0 decays giving the charged mode (6.3*b*), has recently been obtained by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956) in their investigation of strange particle production in $\pi^- + P$ interactions (§ 3.2). For each $\theta_1{}^0$ decay observed whose kinematics corresponded to production through the process $\pi^- + P \rightarrow \Lambda^0 + \theta^0$, a search was made for the decay of the associated Λ^0 particle (whose direction and velocity are then known). Failure to observe this Λ^0 decay in the chamber may be due to escape of the Λ^0 particle before $\Lambda^0 \rightarrow P + \pi^-$ decay or to the occurrence of an unobserved Λ^0 decay such as (6.3*a*). With the known Λ^0 lifetime, the probability of obtaining the observed configuration of Λ^0 decays accompanying all of these $\theta_1{}^0$ decay events was calculated as function of the parameter α_Λ; the most probable value for α_Λ corresponds to the maximum of this probability function, leading to the estimate

$$\alpha_\Lambda = 0{\cdot}3^{+0{\cdot}15}_{-0{\cdot}12}. \qquad \ldots\ldots(6.4)$$

It is of interest to note that an independent lower limit on α_Λ can be obtained from the data of Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp (1956) on $K^- + P$ capture (§ 4.7), with the assumption of charge independence. The total number of $\Sigma^0 + \pi^0$ production events in this experiment is equal to $(14 \pm 2)/\alpha_\Lambda$. From the relative probabilities (2.31) given by charge independence, it follows that this number cannot exceed $\frac{1}{2}[n(\Sigma^+) + n(\Sigma^-)] = 38{\cdot}5$, whence $\alpha_\Lambda \geqslant 0{\cdot}36(\pm 0{\cdot}1)$. (*Note added in proof*: With improved statistics on $\Lambda^0 + \theta^0$ and $\Sigma^0 + \theta^0$ production (168 events where both Λ^0 and $\theta_1{}^0$ decay are seen, 109 events with only $\theta_1{}^0$ decay observed) an improved value of

$$\alpha_\Lambda = 0{\cdot}65 \pm 0{\cdot}05 \qquad \ldots\ldots(6.4a)$$

has recently been reported (Schwartz 1957). Five photons were also observed in association with $\theta_1{}^0$ decay events, which are interpreted as due to the existence of the neutral Λ^0 mode (6.3*a*); these correspond to a probability $0{\cdot}18 \pm 0{\cdot}09$ for the neutral mode.)

The spin and parity of the Λ^0 particle have not yet been clearly established. However, Ruderman and Karplus (1956) have been able to show from analysis of the decay of Λ^0 particles bound in nuclei that the pion emitted in Λ^0 decay most probably carries orbital angular momentum $l = 0$, though $l = 1$ could still be possible. This argument will be discussed in § 6.2. Consequently, only spin-parity values $\frac{1}{2}^-$, and possibly $\frac{1}{2}^+$ and $\frac{3}{2}^+$, need be considered for the Λ^0 particle. For spin $\frac{3}{2}^+$, angular correlations between the Λ^0 decay and its direction and plane of production would generally be expected, for example in the $\pi^- + P$ reaction; this has been discussed (cf. § 3.1) especially by Adair (1955). The present data of Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956) give the best fit with spin $\frac{1}{2}$ according to Adair's criterion, but are not yet sufficient to exclude spin $\frac{3}{2}$. (Spin $\frac{3}{2}$ and higher spin values are now quite strongly excluded as a result of greatly improved statistics in this experiment—see footnote on p. 187.)

Decay of the Σ particles. The first clear estimate of the Σ^- lifetime was obtained by Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956), who found

$1{\cdot}4^{+1{\cdot}6}_{-0{\cdot}5} \times 10^{-10}$ sec on the basis of 14 Σ^- particles of well-established charge. With 49 Σ^- particles arising from the K^-–proton capture reaction (4.31) in the Berkeley 10-in. hydrogen bubble chamber, Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp (1956) have confirmed this result, giving

$$T(\Sigma^-) = 1{\cdot}86(\pm 0{\cdot}26) \times 10^{-10}\ \text{sec.} \qquad \ldots\ldots(6.5)$$

The most significant estimate of the Σ^+ lifetime has also been obtained by Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp (1956), namely

$$T(\Sigma^+) = 0{\cdot}86(\pm 0{\cdot}17) \times 10^{-10}\ \text{sec} \qquad \ldots\ldots(6.6)$$

based on 28 Σ^+ decays from K^-–proton capture events. This is in agreement with an earlier estimate of $0{\cdot}76(\pm 0{\cdot}2) \times 10^{-10}$ sec due to Fry, Schneps, Snow and Swami (1955 a), who included all Σ^+ decays occurring at the end of their range and all $\Sigma^\pm$ decays in flight, the latter being assumed predominantly Σ^+ events. The branching ratio $(P+\pi^0)/(N+\pi^+)$ between the two modes of Σ^+ decay was found to be 14/14 in the data of Solmitz *et al.* For Σ^+ decays occurring at rest,[†] Fry *et al.* found a branching ratio 6/6.

The most detailed evidence[‡] on the existence of the Σ^0 particle has been obtained from the study of Λ^0 decays following the K^-–proton capture reaction; this has been described in § 4.7. Evidence which requires the existence of an exchange reaction $\Sigma^- + P \rightarrow \Sigma^0 + N$ for Σ^- particles coming to rest in a hydrogen bubble chamber will be discussed in § 6.3; from this, it is possible to place limits on the mass difference

$$23\ \text{Mev} > m(\Sigma^-) - m(\Sigma^0) > 2\ \text{Mev} \qquad \ldots\ldots(6.7)$$

with the conclusion that the Σ^0 particle is at least $4m_e$ lighter than the Σ^- particle. The Σ^+ particle was already known to be about $16m_e$ lighter than the Σ^- particle. These mass differences are presumably a consequence of electromagnetic interactions and some preliminary considerations on this question have recently been discussed by Sudarshan and Marshak (1956).

Isotopic spin changes in hyperon decay. As remarked in §§ 4.4 and 5.1, the $\Delta T = \frac{1}{2}$ rule suggested by Gell-Mann for weak decays leading to pions and nucleons has been found to provide a simple interpretation of several striking features of τ and θ decay. Assuming charge independence to hold for the strong interaction between K meson, hyperon and nucleon, it is natural to expect this rule to be effective also in hyperon decay, since hyperon decay and K decay are linked together through this strong interaction.

For Λ^0 decay, this rule requires $T = \frac{1}{2}$ for the final pion and nucleon. In this isotopic-spin state, according to (2.6), the $(\pi^- + P)$ configuration is present with probability 2/3, leading to the value $\alpha_\Lambda = 2/3$. This is in good agreement with the value (6.4*a*) reported recently by the Columbia bubble chamber group.

More generally, the final pion–nucleon state may be a superposition of states of total isotopic-spin $T = \frac{1}{2}$ and $T = \frac{3}{2}$,

$$R_1 \exp(i\delta_1)\left(\sqrt{\tfrac{2}{3}}\,u_P U_- - \sqrt{\tfrac{1}{3}}\,u_N U_0\right) + R_3 \exp(i\delta_3)\left(\sqrt{\tfrac{1}{3}}\,u_P U_- + \sqrt{\tfrac{2}{3}}\,u_N U_0\right). \qquad \ldots\ldots(6.8)$$

[†] *Note added in proof*: This branching ratio has now been refined to $0{\cdot}88 \pm 0{\cdot}12$ from recent data obtained for Σ^+ hyperons resulting from K^- capture in emulsions (Snow 1957).

[‡] See however the recent evidence described in the footnote on page 186, which gives a mass difference (6.7) of $8{\cdot}0 \pm 4{\cdot}5$ Mev.

Takeda (1956) has shown that considerations of time-reversal invariance require R_1 and R_3 to be real amplitudes, the phases δ_1, δ_3 being equal to the phase shifts in the pion–nucleon scattering states of the same angular momentum j_Λ and parity w_Λ as the initial Λ^0 hyperon. For this state (6.8)

$$\alpha_\Lambda = (2R_1^2 + 2\sqrt{2}R_1R_3\cos(\delta_1 - \delta_3) + R_3^2)/3(R_1^2 + R_3^2). \qquad \ldots\ldots(6.9)$$

For each (j_Λ, w_Λ) allowed by Ruderman and Karplus, these phases δ_1, δ_3 are known, so that the amplitudes R_1, R_3 can be deduced from the Λ^0 lifetime and the quantity α_Λ. The most natural interpretation of the value (6.4*a*) for α_Λ is that $R_1 \gg R_3$.

For $\Sigma^\pm$ decay, the $\Delta T = \frac{1}{2}$ rule has less direct implications (Gatto 1956 a, d'Espagnat and Prentki 1956 b, Kawaguchi and Nishijima 1956 b). In Σ^+ decay, this rule involves no restriction, since it allows both $T = \frac{1}{2}$ and $T = \frac{3}{2}$ final pion–nucleon states in arbitrary proportion. Using the wave functions (2.6), the final state may be written in a form analogous to (6.8), the coefficients R_1, R_3 being replaced by amplitudes S_1, S_3 for Σ decay, the phases δ_1, δ_3 being evaluated for the pion–nucleon scattering state (J_Σ, w_Σ) at the energy appropriate to Σ decay. Then α_Σ, the fraction of Σ^+ decays giving $N + \pi^+$, is also given by (6.9) after these replacements have been made. Again the separate amplitudes and their ratio $x = S_1/S_3$ may be deduced from the Σ^+ lifetime and the quantity α_Σ.

The Σ^- decay leads necessarily to the $T = \frac{3}{2}$ state, the $N + \pi^-$ system. With the $\Delta T = \frac{1}{2}$ rule, it may be shown that a relationship exists between this transition and the Σ^+ transitions, namely that this Σ^- amplitude is necessarily $\sqrt{3}S_3\exp(i\delta_3)$. The Σ^- lifetime is then $(1 + x^2)/3$ times the Σ^+ lifetime. Since the phase difference $\delta_1 - \delta_3$ is necessarily small (at most 25°), the value of x is not sensitive to the Σ spin–parity assumed. The present value α_Σ leads to a ratio of order 10 for Σ^-/Σ^+ lifetimes, somewhat larger than the observed ratio. However, owing to the uncertainties in the value of α_Σ and in these lifetimes, this discrepancy cannot be regarded as significant at present, especially as small admixtures of transitions violating the $\Delta T = \frac{1}{2}$ rule may modify appreciably the expression (6.9) for α_Σ.

6.2. *The Nuclear Interactions of the Λ^0 Hyperon*

Hypernuclear fragments. Our knowledge of the Λ^0–nucleon interaction has come primarily from observations on the decay of 'hypernuclear fragments' consisting of a Λ^0 hyperon bound to an ordinary nucleus. The existence of these hypernuclear systems, stable for times comparable with the free Λ^0 lifetime (a very long time on the nuclear scale) was first suggested by Danysz and Pniewski (1953). These authors made this suggestion as a possible explanation for a fragment (charge $Z \sim 4$–6) emitted from a cosmic-ray star, which came to rest after a flight time 4×10^{-12} sec and gave rise to a secondary star with four prongs and a visible energy release of the order of 125 MEV, much too large for any normal nuclear excitation with such a long lifetime.

The first clearly identified hypernucleus was the hypertriton ${}_\Lambda H^3$ consisting of a Λ^0 hyperon bound to a deuteron. This was established by Bonetti, Levi-Setti, Panetti, Scarsi and Tomasini (1954), who observed a fragment with charge $Z = 1$ which decayed at the end of its range to give two particles, a π^- meson and a helium nucleus, whose tracks were collinear. Only for the ^{3}He isotope was the range observed for the helium nucleus compatible with the pion momentum, so that this event could be identified as

$$ {}_{\Lambda}\mathrm{H}^3 \rightarrow {}^3\mathrm{He} + \pi^- + 41{\cdot}7(\pm 1{\cdot}1)\ \text{Mev}. \qquad \ldots\ldots(6.10a)$$

The Q value obtained was comparable with (but not equal to) the value $Q_\Lambda = 36{\cdot}9$ Mev for the Λ^0 decay. The difference between these values may be attributed to the difference $B_{\mathrm{P}} - B_\Lambda$ between the binding energies of the Λ^0 particle in the initial system and of the proton in the final system. From the known value of B_{P} for ^{3}He it follows that $B_\Lambda({}_{\Lambda}\mathrm{H}^3) = 0{\cdot}7(\pm 1{\cdot}1)$ Mev for this event. About ten examples of ${}_{\Lambda}\mathrm{H}^3$ decay have now been reported (cf. table 7), including several decay modes alternative to (6.10*a*),

$${}_{\Lambda}\mathrm{H}^3 \begin{cases} \rightarrow \mathrm{P} + {}^2\mathrm{H} + \pi^-, & \ldots\ldots(6.10b) \\ \rightarrow \mathrm{P} + \mathrm{P} + \mathrm{N} + \pi^-. & \ldots\ldots(6.10c) \end{cases}$$

In general the binding energy B_Λ of the Λ^0 particle is given by

$$B_\Lambda = (m_\Lambda c^2 + m_{\mathrm{i}} c^2) - (\Sigma_f m_f c^2 + Q), \qquad \ldots\ldots(6.11)$$

where $m_\Lambda, m_{\mathrm{i}}$ denote the total rest energy of the Λ^0 particle and of the nucleus to which it is bound, and m_f denotes the mass of a final nuclear fragment (or pion), the sum f being over all fragments (including the pion) of the final system. Since B_Λ represents a small difference between large energy values, the B_Λ values given in table 7 for ${}_{\Lambda}\mathrm{H}^3$ hypernuclei straggle quite widely, mainly as a result of the straggling in the pion range.

A number of other light hypernuclei have also become well established; their decay modes and B_Λ values are also listed in table 7. For these, both 'mesonic' and 'non-mesonic' decay modes are known. Thus, for ${}_{\Lambda}\mathrm{He}^4$, Seeman, Shapiro and Stiller (1955) have reported a non-mesonic decay

$${}_{\Lambda}\mathrm{He}^4 \rightarrow {}^2\mathrm{H} + \mathrm{P} + \mathrm{N} \qquad \ldots\ldots(6.12a)$$

with $B_\Lambda = 1{\cdot}9 \pm 2{\cdot}9$ Mev, whilst Fry, Schneps and Swami (1956 a) have observed two clear examples of the mesonic decay mode

$${}_{\Lambda}\mathrm{He}^4 \rightarrow {}^3\mathrm{He} + \mathrm{P} + \pi^- \qquad \ldots\ldots(6.12b)$$

for which B_Λ values of $1{\cdot}8 \pm 0{\cdot}6$ and $0{\cdot}0 \pm 2{\cdot}0$ Mev were obtained.

Note added in proof: Levi-Setti, Slater, and Telegdi (1957) have recently re-analysed all the mesonic hypernuclear decay events available, including many new events which they have obtained in emulsions exposed to a high energy π^- beam. A single range–energy relation was used for all events and standardized criteria were laid down for estimation of weighting factors for the assignment of errors. Using $Q_\Lambda = 36{\cdot}9 \pm 0{\cdot}2$ Mev, they have obtained the following mean B_Λ values: ${}_{\Lambda}\mathrm{H}^3$, $0{\cdot}02 \pm 0{\cdot}24$ Mev; ${}_{\Lambda}\mathrm{H}^4$, $1{\cdot}44 \pm 0{\cdot}20$ Mev; ${}_{\Lambda}\mathrm{He}^4$, $1{\cdot}70 \pm 0{\cdot}24$ Mev; ${}_{\Lambda}\mathrm{He}^5$, $2{\cdot}56 \pm 0{\cdot}17$ Mev.

The production and decay of hypernuclei. Fry, Schneps and Swami (1955 c, 1956 a) have reported the most extensive investigations of hypernuclear fragments to date. Stars produced in emulsion by cosmic rays, by proton and π^- bombardments from the Cosmotron and Bevatron and by nuclear capture of K$^-$ mesons at the end of their range were examined, and about 100 hyperfragments were observed.

Table 7. Binding Energies for Well-identified Hypernuclei

Hypernucleus	Decay Mode	B_Λ (MEV)	Reference
$_\Lambda H^3$	$\pi^- + {}^3He$	0·7 ± 1·1	Bonetti, Levi-Setti, Panetti, Scarsi and Tomasini (1954)
		0·3 ± 4·0	Debenedetti, Garelli, Tallone and Vigone (1954 a)
		0·8 ± 1·0	Baldo, Belliboni, Ceccarelli, Grilli, Sechi, Vitale and Zorn (1955)
		1·7 ± 4·5	Castagnoli, Cortini and Franzinetti (1955)
	$\pi^- + P + {}^2H$	−1·0 ± 1·2	Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard (1956 b)
		0·5 ± 0·6	Fry, Schneps and Swami (1956 a)
		−0·5 ± 0·6	
		−0·6 ± 0·8	Anderson, Lawlor and Nevin (1955 a)
		1·4 ± 1·0	Skjeggestad and Sorenson (1956)
	$\pi^- + P + P + N$	0·4 ± 0·7	Fry, Schneps and Swami (1956 a)
		1·4 ± 0·6	Haskin, Bowen, Glasser and Schein (1956)
		3·3 ± 2·0	
		−0·9 ± 2·0	
	Weighted Mean Value	0·4 ± 0·3 MEV	
$_\Lambda H^4$	$\pi^- + {}^4He$	1·2 ± 0·8	Friedlander, Keefe and Menon (1955)
		2·2 ± 1·0	Haugerud and Sorenson (1955)
		1·6 ± 3·0	Hornbostel and Salant (1956)
		1·4 ± 0·9	Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard (1956 b)
		0·4 ± 2·4	
		3·3 ± 1·0	Schein, Haskin and Leenov (1955)
	$\pi^- + N + {}^3He$	1·2 ± 0·9	Ladu, Levi-Setti and Scarsi (1956)
		1·2 ± 0·7	Rosselet, Weill and Gailloud (1956)
		5·0 ± 3·0	Hornbostel and Salant (1956)
	$\pi^- + P + N + {}^2H$	1·9 ± 2·0	Fry, Schneps and Swami (1956 a)
	Weighted Mean Value	1·65 ± 0·4 MEV	
$_\Lambda He^4$	$\pi^- + P + {}^3He$	1·8 ± 0·6	Fry, Schneps and Swami (1956 a)
		0·0 ± 2·0	
		1·5 ± 0·6	Gilbert, Violet and White (1956 a)
	$\pi^- + P + P + {}^2H$	1·9 ± 0·6	Fowler and Hansen (1956)
	$P + N + {}^2H$	1·9 ± 2·9	Seeman, Shapiro and Stiller (1955)
	Weighted Mean Value	1·6 ± 0·4 MEV	
$_\Lambda He^5$	$\pi^- + P + {}^4He$	1·3 ± 1·0	Crussard, Fouche, Henessy, Kayas, Leprince-Ringuet, Morellet and Renard (1956 b)
		2·0 ± 0·6	Fry, Schneps and Swami (1956 a)
		2·9 ± 0·5	Fowler and Hansen (1956)
		2·5 ± 1·3	Hill, Salant, Widgoff, Osborne, Pevsner, Ritson, Crussard and Walker (1956)
	Weighted Mean Value	2·2 ± 0·4 MEV	
$_\Lambda Li^7$	$\pi^- + {}^7Be$	4·4 ± 0·7	Gilbert, Violet and White (1956 a)
	$\pi^- + P + {}^2H + {}^4He$	4·4 ± 1·0	Sorensen and Skjeggestad (to be published)
$_\Lambda Be^9$	$\pi^- + P + {}^8Be$	6·5 ± 0·6	Fry, Schneps and Swami (1956 a)

It was found that hypernuclear fragments were emitted from about 5% of K$^-$ capture stars and from about 0·1% of the stars resulting from a high-energy nuclear collision. The relatively high frequency for K$^-$ stars appears natural since we have seen that Λ^0 particles of moderate energies are produced in about 80% of these stars, whereas strange-particle production amounts to only 2 or 3% of the total cross section for high-energy proton or pion interactions. The visible energy release was less than the maximum of $m_\pi c^2 + Q_\Lambda = 176$ MeV allowed for Λ^0 decay for all but one of these fragments. This anomalous case cannot be due to a bound Λ^0 hyperon and will be discussed separately in § 8. Another anomalous fragment of this type has also been reported by Castagnoli, Cortini and Franzinetti (1955) and it is, of course, possible that a small proportion of the events compatible with Λ^0 hyperfragments may also belong to this class of anomalous events. However, all the hyperfragments for which a clear identification has been possible have been established as Λ^0 hyperfragments.

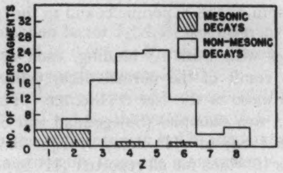


Figure 27. The charge distribution for hyperfragments observed by Fry, Schneps and Swami (1956 a). The average charge Z is between 4 and 5.

The charge Z for each of these fragments is readily obtained if Z is assumed to equal the total charge of all visible distintegration products; of course, this may actually be an under-estimate if these products include a heavy recoil fragment with energy too low to produce a visible track. The charge distribution observed is shown in figure 27. This distribution may be distorted by some bias against finding large Z values for hyperfragments, both for the reason just mentioned and because heavy hyperfragments are more quickly slowed down and may have so short a range as to be essentially unobservable in emulsion. It will be seen that the majority of hyperfragments have $Z \geqslant 3$ and that these generally undergo non-mesonic decay (in fact, only two mesonic decay events were recorded amongst these 88 hyperfragments). Very few of these heavier hypernuclei could be clearly identified. Owing to the large energy released (176 MeV), non-mesonic decay of heavy hypernuclei generally leads to quite a considerable number of nuclear fragments whose masses and charges are often difficult to determine owing to their short range; also the emission of one or two neutrons may generally be possible, which leads to further ambiguity in the interpretation of the decay. The only heavy hypernuclei identified with certainty have involved simple mesonic decays, such as the $_\Lambda\mathrm{Be}^9$ event reported by Fry, Schneps and Swami (1956 a),

$$_\Lambda\mathrm{Be}^9 \rightarrow \pi^- + \mathrm{P} + {}^8\mathrm{Be}, \quad {}^8\mathrm{Be} \rightarrow {}^4\mathrm{He} + {}^4\mathrm{He}, \qquad \ldots\ldots(6.13)$$

for which a value $B_\Lambda = 6{\cdot}5 \pm 0{\cdot}6$ MeV was obtained. Even for helium hypernuclei, the assignment of non-mesonic decay events to $_\Lambda\mathrm{He}^4$, $_\Lambda\mathrm{He}^5$ or $_\Lambda\mathrm{He}^6$ has not generally been possible, apart from the one event (6.12*a*), although about half a dozen non-mesonic decays have been recorded. Also it should be added that, although it is now known that a considerable proportion ($\sim 35\%$) of Λ^0 particles undergo decay to $N+\pi^0$, all mesonic decays identified for hypernuclei have led to a negative pion; this is clearly due to the difficulty of establishing a decay process such as

$${}_{\Lambda}H^3 \rightarrow {}^3H + \pi^0, \qquad \ldots\ldots(6.14)$$

which would be readily confused with a nuclear scattering unless examined carefully, or a process of the type

$${}_{\Lambda}He^4 \rightarrow {}^3He + N + \pi^0, \qquad \ldots\ldots(6.15)$$

which would be the analogue of the relatively probable decay mode (6.12*b*). It will be apparent that there exists quite a strong bias against identification of many decay modes of a given hypernucleus so that considerable caution is needed before conclusions may be drawn concerning their relative frequencies. Gatto (1956 a) has pointed out that there would be considerable interest in these relative frequencies as a guide to the isotopic spin of the hypernuclei and to the isotopic-spin character of the Λ^0 decay process.

For light hypernuclei with weak Λ^0 binding, such as ${}_{\Lambda}H^3$, the decay of the hypernucleus will be a result of the normal decay processes of the Λ^0 particle, the lifetime being rather close to the free Λ^0 lifetime. This is consistent with the evidence on ${}_{\Lambda}H^3$ decay; two examples (Skjeggestad and Sorenson 1956, Brisbout, Friedlander and Iredale 1956) of ${}_{\Lambda}H^3$ decay-in-flight have been observed for a total flight time of $7{\cdot}33 \times 10^{-10}$ sec for all reported ${}_{\Lambda}H^3$ hypernuclei (cf. Friedlander 1957).

However, for heavier hypernuclei, it is apparent that the Λ^0 decay is considerably affected by the presence of other nucleons, non-mesonic decay being the rule for $Z \geqslant 3$. Of course, it is possible for the pion resulting from the decay of a Λ^0 particle within nuclear matter to be absorbed by a pair of nucleons before escaping from the nucleus. However, the nuclei relevant for the observed hypernuclei ($Z \leqslant 8$) are quite small and, making use of known data on the nuclear absorption of pions, Fowler (1956) has shown that no more than 10% of pions would fail to escape, even from carbon or oxygen hypernuclei. The observed non-mesonic decay must therefore be due to an additional mechanism of de-excitation of the Λ^0 particle. This consists of the release of the full Λ^0–nucleon mass difference into nucleon kinetic energy by a virtual Λ^0 decay process, the virtual pion being absorbed directly by one neighbouring nucleon, as suggested by Primakoff and Cheston (1953), thus

$$\begin{aligned} &\Lambda^0 + P \rightarrow (P+\pi^-) + P \rightarrow P + N, \\ &\Lambda^0 + N(P) \rightarrow (N+\pi^0) + N(P) \rightarrow N + N(P). \qquad \ldots\ldots(6.16) \end{aligned}$$

Whereas the pion emitted in the free decay

$$\Lambda^0 \rightarrow P + \pi^- \qquad \ldots\ldots(6.17)$$

has a momentum of about 95 Mev/*c*, the virtual pion involved in the process (6.16) must convey a high momentum ($\simeq 420$ Mev/*c*) to each nucleon, since the final nucleons each receive about 88 Mev kinetic energy. This interaction can therefore be effective only when the Λ^0 particle and nucleon come very close together, within a distance of order $0{\cdot}5 \times 10^{-13}$ cm. In hypernuclei of low *Z*, where the lightly bound Λ^0 particle is well outside the core nucleus for a considerable fraction of the time, this non-mesonic capture process (6.16) can take place relatively infrequently. For the large hypernuclei, where B_Λ is moderately large (e.g. 6·5 MeV for ${}_\Lambda\text{Be}^9$) and the Λ^0 particle generally lies within the core nucleus, this process may be expected to be relatively probable. This is especially so, since the rate of normal Λ^0 decay will be considerably reduced for a Λ^0 particle within nuclear matter by the action of the Pauli principle. For Λ^0 decay from rest, the recoil nucleon has only about 95 MeV/c momentum; since the nucleon states within a large nucleus are already filled for momenta less than about 200 MeV/c, it is clear that the decay of a Λ^0 particle can proceed within a nucleus only in virtue of its motion and that only a small proportion of the final nucleon states energetically possible are allowed by the exclusion principle. We will denote the proportion of final states allowed by $I_\Lambda(A,Z)$, so that the decay probability per unit time is

$$P_{\text{mesonic decay}} = \frac{1}{[T(\Lambda^0)]_{\text{free}}} \times I_\Lambda(A,Z). \qquad \ldots\ldots(6.18)$$

An estimate of this suppression factor $I_\Lambda(A,Z)$ has been given by Primakoff (1956), with values ranging from 0·2 to 0·005 as Z increases from 3 to 10; a mean value 0·05 will be taken below for hyperfragments $Z \geqslant 3$.

The matrix elements for these competing processes (6.16) and (6.17) each depend on the amplitude $M_k(\Lambda^0 \to P + \pi^-)$ for the interaction (6.17) involving a virtual pion of momentum k. Since the momentum $k_c \simeq 420$ MeV/c appropriate to the capture processes (6.16) differs considerably from the momentum $k_d \simeq 95$ MeV/c of the decay process (6.17), the probability of the capture process relative to the decay process will depend directly on the momentum dependence of M_k from $k = k_d$ to $k = k_c$. If the effective radius of the Λ^0 particle is sufficiently small (less than about $0{\cdot}5 \times 10^{-13}$ cm) then, as pointed out by Ruderman and Karplus (1956), the outgoing pion will face a strong centrifugal barrier as a result of its orbital angular momentum l. An expression for the penetration amplitude for such a barrier, valid under the present conditions, has already been given in equation (4.12). This penetration amplitude has a strong momentum dependence, proportional to k^l and it is reasonable to expect that the momentum dependence of M_k will be dominated by this factor, so that

$$M_k = C_l k^l, \qquad \ldots\ldots(6.19)$$

where the coefficient C_l has at most a weak dependence on momentum and on l. On this basis, the ratio

$$R_l = \frac{P\,(\text{non-mesonic capture})}{P\,(\text{mesonic decay})} \qquad \ldots\ldots(6.20)$$

expected for hypernucleus decay, corresponding to the emission of a pion with orbital angular momentum l in (6.16), will have the form

$$R_l = D\left(\frac{k_c}{k_d}\right)^{2l} \simeq (17)^l D, \qquad \ldots\ldots(6.21)$$

where the coefficient D is independent of l. The actual value of D will depend on the interaction of the pion with the nucleon in (6.16), and on the density of nucleons available for the capture process in the neighbourhood of a Λ^0 particle within nuclear matter ; each of these factors can now be estimated with a fair degree of reliability. This quantity R_l represents essentially the internal conversion coefficient for Λ^0 decay in the presence of nuclear matter, and is quite analogous to the internal conversion coefficient for the ejection of orbital electrons from an atom whose nucleus is undergoing decay from an excited state. According to (6.21), this internal conversion coefficient R_l has a rather sensitive dependence on the orbital angular momentum l carried by the outgoing pion in Λ^0 decay.

Ruderman and Karplus (1956) have estimated the coefficient D for two cases :

(*a*) For hypernuclei $Z \geqslant 3$, where the Λ^0 particle is essentially confined to the nuclear volume. Accepting Primakoff's estimate of $I_\Lambda(A,Z)$ and the experimental value of α_Λ (cf. § 6.1), the predicted ratio $R_l(\pi^-)$, when only π^- mesonic decays are counted, becomes

$$R_l(\pi^-) \simeq 40 \times (17)^l. \qquad \ldots\ldots(6.22)$$

The empirical situation is that, for hyperfragments $Z \geqslant 3$, two mesonic (π^-) decays were observed by Fry, Schneps and Swami (1956) for 86 non-mesonic decays.

(*b*) For light hypernuclei $Z \leqslant 2$, where the wave function of the Λ^0 particle extends over a volume large compared with the nuclear volume, so that the effect of the exclusion principle is negligible and $I_\Lambda \simeq 1$. For this case, the value of $R_l(\pi^-)$ naturally depends on the binding energy B_Λ assumed, since the Λ^0 capture rate will be proportional to the probability of its being within the nuclear volume ; the result is

$$R_l(\pi^-) \simeq 0{\cdot}6\sqrt{(B_\Lambda \text{ Mev})} \times (17)^l. \qquad \ldots\ldots(6.23)$$

The observed ratio for helium hypernuclei is close to 1, whilst the values reported for B_Λ are in the range 1 to 3 Mev.

In both cases the data are in good accord with $l = 0$, although it is conceivable that the uncertainties of the calculation and of the experimental results may be such that $l = 1$ should not be completely excluded. From this argument, Ruderman and Karplus have concluded that the Λ^0 particle most probably has spin–parity $\frac{1}{2}^-$, although $\frac{1}{2}^+$ and $\frac{3}{2}^+$ may also be possible.

This conclusion was also reached independently by Nishijima (1955 b) from less detailed arguments based on similar considerations.

Λ^0 *binding energies for hypernuclei and the* Λ^0*–nucleon interaction.* In table 7 we have listed the well-identified hypernuclear decay events reported in the literature, from which have come the best estimates at present available for the Λ^0 binding energies in these hypernuclei. The absence of any reliable evidence for the hyperdeuteron ${}_\Lambda H^2$ or the hyperhelium isotope ${}_\Lambda He^3$ is noteworthy as an indication that the attraction of a Λ^0 particle to one or two protons is considerably weaker than their attraction for a neutron. Similarly, the data on ${}_\Lambda H^3$, ${}_\Lambda He^4$ and ${}_\Lambda Li^7$ show that the binding of a Λ^0 particle to a deuteron, to ${}^3He$ or to ${}^6Li$ is weaker than the binding of a neutron to these systems, so that the Λ^0–nucleon force appears less effective in binding than is the nucleon–nucleon force.

On the other hand, the data on ${}_\Lambda H^4$, ${}_\Lambda He^5$ and ${}_\Lambda Be^9$ show that the Λ^0 binding to ${}^3H$, to ${}^4He$ or to ${}^8Be$ is stronger than the binding of a neutron to these systems, the neutron binding energy B_N being zero for the systems ${}^4H$ and ${}^5He$ and only 1·67 Mev for ${}^9Be$. However, this situation is not inconsistent with the conclusion immediately above but is a consequence of the fact that the restrictions of the exclusion principle which cause B_N to be zero or small for these particular nuclei do not apply to the Λ^0 particle, a particle which is quite distinct from the nucleons. As a result, the Λ^0 particle can occupy the lowest s-orbit in the potential well arising from its attraction by the A nucleons of the nucleus to which it is attached. This potential well will increase in volume as A increases, its depth remaining constant since the density of nucleons in nuclei is sensibly constant with A. Consequently, the Λ^0 binding energy B_Λ may be expected to increase monotonically with increasing A to some limiting value for very large hypernuclei. The present evidence on these binding energies, as shown in figure 28, is in accord with this expectation.

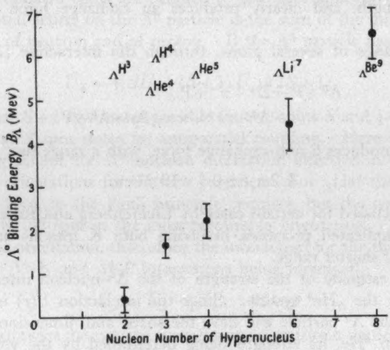


Figure 28. The binding energy B_Λ as function of the number of nucleons, for the well-identified hypernuclei.

The assumption of charge independence for the Λ^0 particle requires that the nuclear interactions in the systems Λ^0+P, Λ^0+N (which then form an isotopic spin doublet) should be identical. This is supported by the equality (within the considerable experimental errors) of the B_Λ values of 1·65 ± 0·4 Mev and 1·60 ± 0·4 Mev for the mirror hypernuclei $_\Lambda H^4$ and $_\Lambda He^4$. The strength of this conclusion becomes more apparent if, as is more appropriate, one compares the potential wells seen by the Λ^0 particle in these two hypernuclei, for the uncertainty ± 0·4 Mev given for each B_Λ corresponds only to ± 5% uncertainty in the well depth. Consequently, we will assume the Λ^0–P and Λ^0–N interactions to be identical; however, they may still depend on whether the Λ^0 and nucleon spins are parallel or antiparallel, and there may also exist spin–orbit coupling in this interaction.

As remarked in § 2.3, charge independence for the Λ^0 particle requires the absence of an interaction (2.20),

$$\Lambda^0 \not\leftrightarrow \Lambda^0+\pi^0. \qquad \ldots\ldots(6.24)$$

Since the interaction of a π^0 meson with a nucleon is known to have opposite sign for N and P (τ_3 coupling), any Λ^0–nucleon interaction which arose from exchange of a single pion through an interaction (6.24) would have opposite sign in the Λ^0–N and the Λ^0–P systems, and this would be sensitively reflected by a difference in the

B_Λ values for ${}_\Lambda H^4$ and ${}_\Lambda He^4$. The $\pm 5\%$ uncertainty in well depth mentioned above allows a coupling parameter for (6.24) at most 1% of the pion–nucleon coupling parameter, all other interactions being assumed charge independent.

Consequently the Λ^0–nucleon interaction must be due to more complicated processes. Contributions to this interaction (Dalitz 1955 b) may come from :

(i) The exchange of a K meson, as discussed in detail by Wentzel (1956 a), for example

$$\Lambda^0 + P \rightarrow P + K^- + P \rightarrow P + \Lambda^0. \qquad \ldots\ldots(6.25)$$

This interaction will have range of order $\hbar/m_K c \simeq 0{\cdot}4 \times 10^{-13}$ cm, the K meson Compton wavelength, and clearly produces an exchange force between Λ^0 and nucleon.

(ii) The exchange of several pions, through the interaction (2.15), giving, for example,

$$\begin{aligned} \Lambda^0 + P &\rightarrow \Sigma^+ + \pi^- + P \\ &\rightarrow \Lambda^0 + (\pi^+ + \pi^- + P) \rightarrow \Lambda^0 + P. \end{aligned} \qquad \ldots\ldots(6.26)$$

This interaction produces a non-exchange force, with a range less than

$$\hbar/2m_\pi c \simeq 0{\cdot}7 \times 10^{-13} \text{ cm}.$$

This has been discussed for certain cases by Lichtenberg and Ross (1956).

(iii) More complicated processes involving both K mesons and pions, giving interactions of still shorter range.

To obtain an estimate of the strength of the Λ^0–nucleon interaction $U(r)$, we may now consider the ${}_\Lambda He^5$ system. Since the interaction $U(r)$ is of short range, the well seen by the Λ^0 particle will have the shape and dimensions of the nucleon distribution $\rho(r)$ in ${}^4He$, its strength being determined by the volume integral of $U(r)$ summed over all the nucleons. Explicitly, this potential will be

$$V(\mathbf{r}) = -\Sigma_{\mathcal{N}} \int U_{\Lambda\mathcal{N}}(\mathbf{r} - \mathbf{r}_{\mathcal{N}})\, \rho(\mathbf{r}_{\mathcal{N}})\, d_3\mathbf{r}_{\mathcal{N}}, \qquad \ldots\ldots(6.27)$$

which may be approximated by

$$\rho(r) \left\{ \Sigma_{\mathcal{N}} \int U_{\Lambda\mathcal{N}}(\mathbf{r} - \mathbf{r}_{\mathcal{N}})\, d_3\mathbf{r}_{\mathcal{N}} \right\} = -\bar{U}_4 \rho(r), \qquad \ldots\ldots(6.28)$$

the density $\rho(r)$ being normalized to unity. Fortunately this function $\rho(r)$ has been determined directly for the alpha particle by electron-scattering experiments (McAllister and Hofstadter 1956) ; the result obtained was that $\rho(r)$ has essentially Gaussian shape with r.m.s. radius $1{\cdot}61 \pm 0{\cdot}05 \times 10^{-13}$ cm. With these parameters and the best value for B_Λ, the Schrödinger equation for the bound Λ^0 particle may be solved to find $\bar{U}_4$; in this way the volume integral†

$$\bar{U}_4 \simeq 1{\cdot}71/m_\pi^2 \qquad \ldots\ldots(6.29)$$

is obtained for the four nucleons of the alpha particle.

It is striking that the B_Λ value for ${}_\Lambda He^5$ appears so little larger than the value for ${}_\Lambda He^4$ despite the increase in the number of nucleons to attract the Λ^0 particle and the more concentrated shape of the ${}^4He$ core. This may be shown quantitatively by calculating $\bar{U}_3$ from the ${}_\Lambda\mathrm{He}^4$ data. There is less certainty here concerning the shape of the ^{3}He core, but taking Gaussian shape again, with r.m.s. radius to fit the observed Coulomb energy of ^{3}He, the following value,

---

† Note that this unit $1/m_\pi^2$ for the volume integral of the interaction is equal to 384×10^{-39} Mev cm^3.

$$\bar{U}_3 \simeq 1{\cdot}91/m_\pi^2 \qquad \ldots\ldots(6.30)$$

is obtained. If the Λ^0–nucleon interaction were spin-independent, a value of $\bar{U}_3$ equal to about $\frac{3}{4}$ of $\bar{U}_4$ would be expected; the fact that (6.30) is considerably larger than this expectation may therefore be attributed to the spin dependence of this interaction.

In ${}_\Lambda\mathrm{He}^5$, the nucleon spins are coupled to give zero spin for the alpha particle, so that the potential acting on the Λ^0 particle is the sum of the interactions for both spin orientations of neutron and of proton. If the Λ^0 particle has spin S, then this sum is given by

$$\bar{U}_4 = 4\{S\bar{U}_a + (S+1)\,\bar{U}_p\}/(2S+1), \qquad \ldots\ldots(6.31)$$

since there are just $S+1$ spin states possible for two spins S and $\frac{1}{2}$ which are coupled to be parallel, and S spin states for antiparallel coupling. Here $\bar{U}_p$ and $\bar{U}_a$ denote the volume integrals for the Λ^0–nucleon interaction effective in these parallel and antiparallel spin orientations for the two particles. For ${}_\Lambda\mathrm{He}^4$ the proton spins are coupled to total zero, as the Pauli principle requires, but the neutron spin is free to couple with the Λ^0 spin in the most favourable orientation. If $\bar{U}_>$ denotes the interaction in this orientation, then since the interaction for the two coupled protons is just $\bar{U}_4/2$ (the Λ^0–N and Λ^0–P interactions being identical),

$$\bar{U}_> = \bar{U}_3 - \tfrac{1}{2}\bar{U}_4 \simeq 1{\cdot}05/m_\pi^2, \qquad \ldots\ldots(6.32)$$

which is more than twice the average over all spin orientations, namely $\bar{U}_4/4 \simeq 0{\cdot}43/m_\pi^2$. This comparison gives some measure of the degree of spin dependence which the present data require in the Λ^0–nucleon interaction. It should be emphasized here that, owing to the short range of this interaction and the fact that the Λ^0 particle is in an s-state, since it is in its ground state in the well provided by the core nucleus, it is only the Λ^0–nucleon interaction for relative s-state motion which is effective in these nuclei.

This conclusion of spin dependence is confirmed by consideration of the ${}_\Lambda\mathrm{H}^3$ system. Neglecting the distortion of the deuteron by the presence of the Λ^0 particle, an estimate of $\bar{U}_2$ may be obtained from the value of B_Λ for ${}_\Lambda\mathrm{H}^3$ and the known (Hulthén) wave function of the deuteron, with the result (Dalitz and Downs 1956, unpublished),

$$\bar{U}_2 \simeq 1{\cdot}70/m_\pi^2, \qquad \ldots\ldots(6.33)$$

which appears to be in reasonable accord with a similar calculation by Derrick (1956) for physically reasonable ranges of the Λ^0–nucleon interaction. This value (6.33) is much larger than one-half of $\bar{U}_4$, the value which would be expected for a spin-independent Λ^0–nucleon interaction. Since the nucleon spins are parallel in the deuteron, the Λ^0 spin can generally orient itself into the energetically most favourable configuration with respect to the spins of both nucleons at the same time, so that a value $\bar{U}_2 \sim 2\bar{U}_> = 2{\cdot}1/m_\pi^2$ is generally to be expected. However, there is an exceptional case where $S = 1/2$ and the antiparallel Λ^0–nucleon spin orientation is the most attractive. For this case $\bar{U}_a = 1{\cdot}05/m_\pi^2$ and $\bar{U}_p = 0{\cdot}22/m_\pi^2$

according to equations (6.29) and (6.32). Since it is not possible for both nucleon spins to be anti-parallel to a Λ^0 spin $S = 1/2$, the value expected for $\bar{U}_2$ in this exceptional case is somewhat smaller, namely $\bar{U}_2 = (3\bar{U}_a + \bar{U}_D)/2 = 1{\cdot}69/m_\pi^2$. Considering the uncertainties of these estimates, either of these values should probably be regarded as compatible with (6.33).

To give a measure of the strength of this interaction (6.32), let us assume that these forces are due to K exchange alone. The Λ^0–nucleon interaction potential would then be, roughly,

$$U_{\Lambda\mathcal{N}} = -g_K{}^2 \mathcal{P}_{\Lambda\mathcal{N}} \frac{\exp(-m_K r)}{r}, \qquad \ldots\ldots(6.34)$$

where $g_K{}^2$ is the coupling parameter for the interaction $\mathcal{N} \longleftrightarrow \Lambda^0 + K$ and $\mathcal{P}_{\Lambda\mathcal{N}}$ denotes the exchange of the spin and space coordinates for Λ^0 and nucleon (here spin $\frac{1}{2}$ has been assumed for the Λ^0 particle). By comparing the volume integral of (6.34) with (6.32), the value $g_K{}^2 \simeq 1{\cdot}0$ is obtained. This value is in accord with the estimate given by Iwao (1955) from variational calculations on the ${}_\Lambda H^3$ system. This result indicates that the Λ^0–nucleon force necessary to account for the binding data for hypernuclei actually corresponds to a very strong interaction, its weakness in binding being attributable rather to its short range.

For the weakly bound ${}_\Lambda H^3$ system, consisting of a Λ^0 particle bound to a deuteron, a more delicate treatment is desirable, since the strong short-range Λ^0–nucleon interactions will cause considerable distortion from the deuteron structure for the neutron and proton. A step in this direction is provided by variational calculations using simple trial wave functions with suitable flexibility, as carried out by Iwao (1955) and Dalitz and Downs (1956, unpublished).

Apart from the one exceptional case (see above), the only Λ^0–nucleon interaction effective in this system is $\bar{U}_>$, corresponding to the most attractive Λ^0–nucleon spin orientation. Owing to the open structure of the system, such calculations will be sensitive primarily to the scattering length $a_0(\Lambda^0, \mathcal{N})$ at zero energy for the Λ^0–nucleon system with this spin orientation. With various assumptions concerning the forces, the observed binding energy B_Λ leads to cross sections of about 1 barn for the Λ^0–nucleon scattering at zero energy in this spin state. In these cases the Λ^0–nucleon interaction deduced was always insufficient to produce a bound Λ^0–nucleon system ;† this conclusion is in accord with the cruder estimate (6.32) for ranges of the Λ^0–nucleon interaction exceeding $\hbar/m_K c$.

In the exceptional case, as mentioned above, the ${}_\Lambda H^3$ binding energy will depend on the scattering lengths for both Λ^0–nucleon spin states so that no clear-cut statement can be made in this way. If the ratio $\bar{U}_a/\bar{U}_D$ is as indicated above for this case, then the Λ^0–nucleon interaction for antiparallel spins would be quite close to the strength needed for binding of a hyperdeuteron if the interaction has range as short as $\hbar/m_K c$. In this case the Λ^0–nucleon cross section could be rather large at low energies and this large scattering would have a strong effect on the energy-sharing between the final particles in a reaction such as (4.35).

In terms of the present considerations the absence of ${}_\Lambda He^3$ can readily be understood. In this system (which would be a member of a $T = 1$ multiplet, ${}_\Lambda He^3$,

---

† It is of interest to note here that Nishijima (1955 b) has shown by an ingenious variational argument that, irrespective of its shape or range, the Λ^0–nucleon interaction must have a strength at least 0·19 of the strength necessary to bind a Λ^0–nucleon system.

${}_{\Lambda}H^3$ and ${}_{\Lambda}n^3$) the protons, having s-wave relative motion, would necessarily have their spins opposed; the interaction seen by the Λ^0 particle would therefore be only the average interaction $\bar{U}_4/2$, very considerably less than the interaction $\bar{U}_2$ effective in the $T = 0$ state ${}_{\Lambda}H^3$. Furthermore, the ${}^1S$ proton–proton interaction is considerably weaker than the ${}^3S$ neutron–proton interaction which is effective in the $T = 0$ ${}_{\Lambda}H^3$ state. The absence of a bound state for the $T = 1$ multiplet containing ${}_{\Lambda}He^3$ therefore appears natural (Jones and Knipp 1955, Dalitz 1955 b) in view of the weakness of the interactions effective in this configuration relative to those in the $T = 0$ configuration.

Finally we consider the spins and parities of these hypernuclear states. The ${}_{\Lambda}H^4$ system has a characteristically energetic decay mode,

$$ {}_{\Lambda}H^4 \rightarrow He^4 + \pi^- + \sim 55 \text{ MeV}, \qquad \ldots\ldots(6.35)$$

leading to two spin-zero particles. If the spin of ${}_{\Lambda}H^4$ is J, then the relative motion of these particles has orbital angular momentum J and the total parity of the system is $-(-1)^J$, including the negative parity of the pion. The spin–parity of ${}_{\Lambda}H^4$ is therefore limited to values $0^-, 1^+, 2^-, 3^+, \ldots$. For any assumed spin–parity of the Λ^0 particle one may then deduce whether $\bar{U}_>$, the interaction effective between the Λ^0 particle and the proton of ${}_{\Lambda}H^4$, corresponds to the parallel or antiparallel spin state. For example, with spin $\frac{1}{2}^-$ for the Λ^0 particle, only spin 0^- is possible for ${}_{\Lambda}H^4$ (since the Λ^0 particle is necessarily in an S-state, its ground state in this system), which requires antiparallel spins to be preferred in the Λ^0–proton system, so that $\bar{U}_a \equiv \bar{U}_>$. The interaction in the parallel spin state may then be deduced from (6.31) and (6.29). One may then go on to deduce the spin–parity of the remaining hypernuclei; ${}_{\Lambda}H^3$ and ${}_{\Lambda}Li^7$ will necessarily have the same spin and parity, namely $\frac{1}{2}^-$ in the present instance, whilst ${}_{\Lambda}He^5$ and ${}_{\Lambda}Be^9$ will always have the spin–parity of the Λ^0 particle. For the spin–parity values allowed by Ruderman and Karplus, these conclusions are collected in table 8.

Table 8. The Spin-dependence of the Λ^0–nucleon Interaction for Various Λ^0 Spin-parity Assumptions; Spin Assignments for ${}_{\Lambda}H^4$, ${}_{\Lambda}H^3$ and ${}_{\Lambda}He^5$ are given for each Case

Spin–parity of Λ^0 particle	$\frac{1}{2}^-$	$\frac{1}{2}^+$	$\frac{3}{2}^+$
Spin–parity of ${}_{\Lambda}H^4$	0^-	1^+	1^+
$m_\pi^2 \bar{U}_p$	0·22	1·05	0·07
$m_\pi^2 \bar{U}_a$	1·05	−1·44	1·05
Spin–parity of ${}_{\Lambda}H^3$ and ${}_{\Lambda}Li^7$	$\frac{1}{2}^-$	$\frac{3}{2}^+$	$\frac{1}{2}^+$
Spin–parity of ${}_{\Lambda}He^5$ and ${}_{\Lambda}Be^9$	$\frac{1}{2}^-$	$\frac{1}{2}^+$	$\frac{3}{2}^+$

These cases may be distinguished in principle by the study of angular correlations in the decay of certain hypernuclei. If angular correlations are found between the decay (6.35) and the process of ${}_{\Lambda}H^4$ production, for example between the pion direction and the initial direction of the ${}_{\Lambda}H^4$ hypernucleus, then the Λ^0 spin will necessarily be $\frac{1}{2}^+$ or $\frac{3}{2}^+$. If angular correlations between the pion direction in the ${}_{\Lambda}H^3$ decay (6.10*a*) and the initial ${}_{\Lambda}H^3$ direction are found to exist, this would indicate spin $\frac{3}{2}^+$ for ${}_{\Lambda}H^3$ and therefore spin $\frac{1}{2}^+$ for the Λ^0 particle. If no angular correlations are found for either of these decays for any production process, this would provide support for the assignment $\frac{1}{2}^-$ for the Λ^0 particle suggested by Ruderman and Karplus (1956). The study of angular correlations in light hypernuclei might be conveniently carried out in a helium bubble chamber, following reactions such as

$$K^- + {}^4\text{He} \begin{cases} \rightarrow {}_\Lambda H^4 + \pi^0, & \ldots\ldots(6.36a) \\ \rightarrow {}_\Lambda H^3 + N. & \ldots\ldots(6.36b) \end{cases}$$

The former reaction (6.36*a*) is of particular interest. If ${}_\Lambda H^4$ has spin $0-$, this reaction is completely forbidden for a scalar K particle, whereas it can occur for a pseudoscalar particle from any partial wave. For spin 1^+, on the other hand, the reaction is forbidden for capture of a pseudoscalar particle from the s-wave, although it is allowed for capture from p or higher states (the angular correlation will be $\sin^2\theta$ for p-wave capture); in this case a scalar particle can also be captured either from the s-wave (giving angular correlation $\cos^2\theta$) or from any higher wave.

The interpretation of this spin dependence for the Λ^0–nucleon interaction is far from clear at present. The K exchange mechanism (6.25) necessarily leads to a strong spin dependence. For example, an s-wave $\mathcal{N} \rightarrow \Lambda^0 + K$ interaction would lead to a Λ^0–nucleon interaction of the form (6.34). Whereas the space exchange operator in this expression may be replaced by unity (only s-wave Λ^0–nucleon states being effective in the light hypernuclei), the spin exchange operator is $+1$ for parallel spins and -1 for antiparallel spins in the Λ^0–nucleon system. This corresponds to a strong spin-dependence, with $\bar{U}_p$ negative and $\bar{U}_a = -\bar{U}_p$, which does not agree with any of the cases listed in table 8. However, with a pseudoscalar $\mathcal{N} \rightarrow \Lambda^0 + K$ coupling, Wentzel (1956) found attraction in both spin states, with $\bar{U}_a = 3\bar{U}_p$; this would be compatible with the first column of table 8 if this interaction is transmitted by a scalar K meson. For the exchange of several pions, as in (6.26), Lichtenberg and Ross (1956) found that with spin 1/2 (either parity) for the Λ^0 particle and the same spin–parity for the Σ particle, parallel spins are favoured, antiparallel spins giving a repulsive interaction. It seems probable that the observed Λ^0–nucleon interaction may result from the exchange of both scalar and pseudoscalar K particles as well as of pions, so that the interpretation of the observed spin-dependence may well be rather complicated. Angular correlation studies in ${}_\Lambda H^3$ and ${}_\Lambda H^4$ decay would provide information very desirable at the present time.

6.3. *The Nuclear Interactions of the Σ Hyperons*

Σ^- capture in hydrogen. Seven examples of Σ^-–proton interactions observed in a hydrogen bubble chamber by Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp (1956) in the course of their study of K^- absorption in hydrogen (cf. § 4.7) have provided our most direct information on the Σ–nucleon interaction. In two of these interactions, a Λ^0 hyperon of about 40 Mev kinetic energy was observed, in accord with the theoretical expectation of the reaction (2.27)

$$\Sigma^- + P \rightarrow \Lambda^0 + N. \qquad \ldots\ldots(6.37)$$

In three cases, no interaction was observed at the end of the Σ^- range; this can again be accounted for by reaction (6.37), since the Λ^0 particle is known to have neutral decay modes. However, in each of the remaining two cases, an associated Λ^0 decay corresponding to a very slow Λ^0 particle was observed, the Λ^0 kinetic energies being 4·5 and 2·8 MeV respectively. These events may be attributed to a charge-exchange process for the Σ^- particle,

$$\Sigma^- + P \rightarrow \Sigma^0 + N, \quad \Sigma^0 \rightarrow \Lambda^0 + \gamma. \qquad \ldots\ldots(6.38)$$

However, this requires that the Σ^0 particle have appreciably smaller mass than the Σ^- particle, since the neutron is already heavier than the proton, and the observed Λ^0 kinetic energies correspond to a mass difference $[m(\Sigma^-) - m(\Sigma^0)]$ between $4m_e$ and $46m_e$.

The observation of the reaction (6.37) implies only that this interaction is stronger than the weak Σ^- decay interactions, for it is conceivable that Σ and Λ particles may still differ in some essential quantum number (Utiyama and Tobocman 1955), causing reaction (6.37) to be slow. However, it is natural to expect the charge–exchange interaction (6.38) to be strong even in this case. Now the phase space available for reaction (6.38) is between 15% and 50% of that for (6.37), so that the observed ratio 2/2 for these processes requires (6.37) also to be a strong interaction. Indeed, since the final state (6.37) has $T = \frac{1}{2}$, only the $T = \frac{1}{2}$ component of the initial system can undergo this reaction, whilst (6.38) is possible for both $T = \frac{1}{2}$ and $T = \frac{3}{2}$ components.

Σ^- *capture in nuclei*. Systematic evidence on Σ^- capture by nuclei may be obtained by following out Σ^- tracks from K^- capture stars in emulsion. Some information obtained in this way has been reported by Chupp, Goldhaber, Goldhaber and Webb (1955), Fry, Schneps, Snow and Swami (1955 a, b), Hornbostel and Salant (1956) and by Haskin, Bowen and Schein (1956). About one-third of these Σ^- tracks end without visible star, only blobs due to energetic Auger electrons being seen. For μ^- capture it is known that such Auger electrons are observed for only about 25% of μ^- endings, so that it appears likely that about 75% of Σ^- tracks ending without visible star are not identified, being indistinguishable from proton tracks.

The Σ^- stars giving visible tracks are generally small stars with one or perhaps two prongs of total energy 10–20 MeV. The emission of a slow hyperfragment is observed from a small proportion (perhaps 10%) of these stars (for example, Schein, Haskin and Leenov 1955, Ceccarelli, Dallaporta, Grilli, Merlin, Salandin, Sechi and Ladu 1955, Rosselet, Weill and Gailloud 1956). However, quite energetic stars are also seen; these are especially prominent in the less systematic cosmic-ray data, where there is a natural bias favouring discovery of large stars. The cosmic-ray data (12 negative hyperon capture events) have been summarized by Ceccarelli, Dallaporta, Grilli, Merlin, Salandin, Sechi and Ladu (1955), and about one-third of these events lead to a visible energy release exceeding 100 MeV.

The theoretical interpretation of this wide range of star sizes has been discussed both by Ceccarelli, Dallaporta, Grilli, Merlin, Salandin, Sechi and Ladu (1955) and by Friedlander, Keefe and Menon (1955). It is natural to interpret these Σ^- capture stars in terms of the reaction (6.37). The charge–exchange reaction (6.38) with its small energy release is likely to be strongly suppressed, partly because considerable energy is absorbed just to change a proton into a neutron in the nuclei present in emulsion and partly because a high proportion of the final neutron states which are energetically permissible are already filled. It should be noted also that Σ^- capture by a neutron is not possible through strong interactions, since there is no negatively charged system containing two baryons and with strangeness $s=+1$ which is energetically possible. There is no negative system corresponding to the final state (6.37) and the energy available is insufficient for an additional pion.

From this initial interaction (6.37), two neutral particles result with a total energy of approximately 80 MeV. There are then various possibilities:

(i) The Λ^0 particle escapes at once from the nucleus with relatively little interaction. This leaves a nucleus with at most 40 MeV excitation energy and probably rather less, as the neutron also has a good chance of escaping. The resulting disintegration star will be small, and often not produce visible prongs.

(ii) The Λ^0 particle may collide with other nucleons, or groups of nucleons, carrying them with it to form a hyperfragment which is able to escape from the nucleus.

(iii) The Λ^0 particle may lose so much energy in collisions that it cannot escape from the nucleus directly. However, it is still possible that the Λ^0 particle may escape through the evaporation processes which occur for a time of order 10^{-20} sec after the initial reaction; this may leave as much as about 80 MeV to other nuclear particles.

(iv) The Λ^0 particle may be captured by the nucleus and not escape during the evaporation period. The system is now a hypernucleus of unusually large Z in which the Λ^0 particle may have quite high binding (using the result (6.29) and the known nucleon density in nuclei, the well depth will be about 17 MeV). In general this heavy hypernucleus will undergo non-mesonic decay releasing an additional 176 MeV, through the capture processes (6.16). This process will give rise to energetic stars and provides a satisfactory explanation of the energetic protons observed from some of the Σ^- stars (Goldsack and Lock 1956).

Σ *hypernuclei.* Owing to the existence of the fast reaction (6.37), a Σ^- particle cannot remain stable in a nucleus containing protons. Consequently the binding of a Σ^- particle to a nucleus cannot provide an explanation of anomalous hyperfragments whose decay releases energy exceeding the maximum of 176 MeV for Λ^0 decay; this remark applies equally to the Σ^+ particle in a nucleus containing neutrons, owing to the charge symmetric reaction

$$\Sigma^+ + \mathrm{N} \rightarrow \Lambda^0 + \mathrm{P}. \qquad \ldots\ldots(6.39)$$

Holladay (quoted by Sachs 1955) has emphasized that, if the Σ^-–neutron interaction were attractive and strong enough to produce a bound state, the Σ^-–neutron hypernucleus would then remain stable for a time comparable with the free Σ^- lifetime. Such an interaction would include a contribution from exchange of a single pion between Σ^- and neutron, arising from

$$\Sigma^\pm \rightarrow \Sigma^\pm + \pi^0, \qquad \ldots\ldots(6.40)$$

which would be important because of its long range relative to other possible processes (cf. of the type (6.25)); however, nothing is known concerning the relative sign of pion–Σ and pion–nucleon couplings and it may be that this interaction would turn out repulsive. With an attractive Σ^-–nucleon force, a bound hypernucleus (Σ^-NN) would also be conceivable; the interaction seen by the Σ^- particle is $(3U_p + U_a)/2$ in this case compared with the greater of U_p and U_a, which is effective in the (Σ^-N) system. If the spin dependence of this interaction is not too strong (or if U_p is the dominant attraction), then this three-particle system would be more likely to form a bound hypernucleus than the (Σ^-N) system (Dalitz 1956 a). Apart from the effect of the Coulomb repulsions, similar remarks hold for the (Σ^+P) and (Σ^+PP) systems. No clear evidence requiring the existence of any of these Σ hypernuclei has yet appeared.

6.4. *The Ξ Hyperons*

Relatively little is known concerning the Ξ particles. The decay mode ($\Lambda^0+\pi^-$) for the Ξ^- particle is well established, the Q value being 65 ± 4 MEV on the basis of seven cosmic-ray events; its lifetime is known only to be of the order 10^{-10} sec.

Isotopic spin $T=\frac{1}{2}$ is assigned to the Ξ hyperons in the Gell-Mann scheme, to correspond with strangeness $s=+2$, following equation (2.23). This requires the existence of a corresponding Ξ^0 particle, for which direct evidence would be difficult to obtain, since its most probable decay modes will be neutral (cf. § 2.3). However, this strangeness assignment also requires that production of a Ξ^- particle in a nucleon–nucleon collision should be accompanied by particles of total strangeness $s=+2$, namely two K particles, a reaction of the type (2.22). A Ξ^- production event which is in accord with this expectation has recently been reported by Sorrels, Leighton and Anderson (1955). Four V events were found in the chamber. Two of these were identified as θ_1^0 decays whose momentum passed through the only nuclear interaction apparent in the chamber. A third V^0 decay, not directly identified, passed through the decay point of a negative V decay for which the $Q(\Lambda^0, \pi^-)$ computed was in accord with the Q value known for the Ξ^- decay; if this be assumed a Ξ^- decay, the Q value of the third V^0 decay is then consistent with the Λ^0 Q value. The event is interpreted as production of two θ_1^0 particles in association with a Ξ^- particle; the primary particle (momentum >4 GEV/c) was not identified.

Gatto (1956 c) has recently suggested that information on the Ξ and Λ spins may be obtained by studying the angular correlation between the directions (specified in the Ξ^- and Λ^0 rest frames respectively) of the first and second pions in the cascade sequence

$$\Xi^- \rightarrow \Lambda^0+\pi^-, \quad \Lambda^0 \rightarrow P+\pi^-. \qquad \ldots\ldots(6.41)$$

However, among the Λ^0 spin values allowed by Ruderman and Karplus (cf. § 6.3), a non-isotropic angular correlation is possible in this process only with spin $\frac{3}{2}^+$ for the Λ^0 particle. No obvious deviation from isotropy exists in the limited data now available, but this is scarcely statistically significant.

None of the negative hyperon stars observed in emulsions has been identified as Ξ^- capture stars; this would require the incident hyperon to be distinguished from a Σ^- particle, presumably by mass measurement, which would appear very difficult. There is every reason to believe that negative hyperon stars following K^- capture are necessarily Σ^- stars, but the cosmic-ray data (Ceccarelli, Dallaporta, Grilli, Merlin, Salandin, Sechi and Ladu 1955) could include some Ξ^- capture stars. A Ξ^-–proton system having strangeness $s=-2$, can undergo a fast reaction to produce two Λ^0 hyperons:

$$\Xi^-+P \rightarrow \Lambda^0+\Lambda^0. \qquad \ldots\ldots(6.42)$$

However, the Ξ^-–neutron system is expected to be stable with respect to fast processes, since all other negatively charged systems of two baryons with $s = +2$ have greater energy. Following nuclear capture of a Ξ^- hyperon by process (6.42), two Λ^0 particles are produced, each with about 14 Mev. The resultant nuclear disintegration can then occur in one of the following ways (Friedlander, Fujimoto, Keefe and Menon 1955):

(i) Both Λ^0 particles may escape directly or through the subsequent evaporation processes. The nuclear excitation energy is at most 28 Mev so that only a very small star will generally result.

(ii) One of the Λ^0 particles escapes whilst the other remains trapped within the nucleus. The residual nucleus is then a large hypernucleus which will undergo non-mesonic decay in due course, releasing about 180 Mev; this will give rise to quite a large star, which could not be distinguished from the large Σ^- stars (cf. § 6.3). It is of course possible that the escaping Λ^0 particle may carry other nucleons with it as a hyperfragment of low energy, which subsequently decays close to the Ξ^- capture event.

(iii) Both Λ^0 particles may be trapped inside the nucleus, the total excitation energy being of order 30 Mev. By neutron emission and gamma radiation this system would reach its ground state, a hypernucleus containing two Λ^0 particles. Such a hypernucleus would be of the greatest interest as it could lead to information concerning the nuclear interaction between Λ^0 particles. Its disintegration would be a two-stage process. The first Λ^0 decay releases about 180 Mev energy; in this disintegration, the second Λ^0 particle may escape by evaporation, either alone or as constituent of a hypernucleus which would then undergo decay close by, or it may remain in the recoil nucleus, which may move an observable distance before the second Λ^0 decay. In this latter case a total energy release of about 350 Mev will occur in this sequence of events, which may be expected to lead frequently to a double-centred star. However, no negative hyperon capture events which require this explanation have yet been reported.

§ 7. Relationships between the K Particle Decay Modes

Quite a number of properties have been established for each of the K particle decay modes discussed in § 4. For the parent particle of each mode a mass value has been obtained both by the momentum–range method and (for $K_{\mu 2}$, $K_{\pi 2}$ and τ) from the energy of the decay products, and these values are all compatible with a common mass value, although the statistical errors quoted would still allow mass differences to be as much as several electron masses. The lifetime measurements described in §§ 4.2 and 4.3 allow the $K^+_{\mu 2}$, $K^+_{\pi 2}$ and τ lifetimes to differ by no more than 20%, and the limited data available on the $K^+_{\mu 3}$ and K^+_{e3} decays (Hoang, Kaplon and Yekutieli 1957) also give lifetimes compatible with these values. Further, the relative frequencies of these decay modes (cf. § 4.6) appear to be unaffected (within statistical uncertainties) by changes in the parameters of the production process (e.g. bombarding energy, angle of emission, etc.) or by nuclear scattering of the K^+ beam. All of these observations are compatible with the simple hypothesis that all of the K^+ decay modes observed represent competing decay schemes for a single parent K^+ particle.

On the other hand, the distribution observed for the pions resulting from τ^+ decay (cf. § 4.3) indicates that their parent particle has a spin and parity such that decay to two pions is forbidden by the requirements of parity conservation. None of the features characteristic of three pions whose total spin and parity are both even or both odd are apparent in these data. At least the $K_{\pi 2}$ and τ modes must result from distinct K particles, differing either in spin or in parity. The simplest hypothesis is that these particles have spin 0^+ and 0^- respectively, but higher even spin values (again with even parity for $K_{\pi 2}$, odd parity for τ particle) are not excluded for either or both of these particles. It is natural to expect that the remaining K decays represent competing decay schemes for one or other of these two particles, and there is at present no evidence contradicting this assumption.

The existence of two distinct K particles whose mass values are at present indistinguishable presents us with many problems. As pointed out by Lee and Yang (1956 a), the existence of such a close doublet must almost certainly reflect some deep-lying symmetry common to the strongly interacting particles, with important consequences for them (cf. § 7.3, below). The most striking puzzle is provided by the close equality of the lifetimes observed for these two K particles, despite the difference between their decay processes. The proposals which have been made to account for this observation will be discussed in the following three sections.

As discussed in § 4.7, the nuclear interactions observed for K^- particles are consistent with Gell-Mann's view that they are the antiparticles to the corresponding K^+ particles. Symmetry between particles and their antiparticles requires that the negative K particles also form a τ^-–θ^- doublet. However, since the K^- decay modes will be simply charge conjugate to the K^+ decay modes, with particle–antiparticle symmetry, data on K^- decay modes can throw no new light on the τ–θ question; in any case, relatively little is known concerning K^- decay. It is possible that nuclear absorption of τ^- and θ^- particles may proceed through different reactions but there is no evidence at present to indicate this (cf. § 7.3). The neutral K particles should also appear as a $\tau-\theta$ doublet. There is some evidence for a $\tau^0 \to 3\pi$ decay event (cf. § 5.3) but very little is known concerning the possibility of alternative decay modes for the τ^0 particles and the lifetime of the τ^0 particles; comparison between the lifetimes for the τ^0 and the θ^0 particles would be especially instructive concerning the origin of the τ^+, θ^+ lifetime equality.

7.1. *Electromagnetic Transitions between* K *Particles?*

The possibility of a parent–daughter relationship between the τ and θ particles was first advanced by Lee and Orear (1955) as a means to account for the lifetime identity observed for K^+ particles. If there is a mass difference between the τ and θ particle, the known selection rules for strong and electromagnetic interactions between strange particles would allow a γ transition from the heavier to the lighter particle, thus

$$\tau^+ \to \theta^+ + \gamma, \qquad \ldots\ldots(7.1a)$$

$$\text{or} \quad \theta^+ \to \tau^+ + \gamma, \qquad \ldots\ldots(7.1b)$$

according as the τ^+ or θ^+ particle is the heavier. Such a γ transition is forbidden, of course, if both τ and θ have zero spin; in this case, however, a transition involving the emission of two photons is still permitted:

$$\tau^+ \rightarrow \theta^+ + \gamma + \gamma, \qquad \text{......(7.2}a\text{)}$$

$$\text{or} \quad \theta^+ \rightarrow \tau^+ + \gamma + \gamma. \qquad \text{......(7.2}b\text{)}$$

If the situation (7.1*a*) or (7.2*a*) holds, for example, this radiative decay of the τ^+ particle will be in competition with the 3π decay and any other decay modes of the τ^+ particle. The secondary θ^+ particle following this radiative decay will then give rise to $K^+_{\pi 2}$ events. The lifetime observed for these $K^+_{\pi 2}$ events would be due almost entirely to the lifetime of the parent τ^+ particle if the lifetime for θ^+ decay were rather shorter than the τ^+ lifetime (say $< 10^{-9}$ sec). In this case any directly produced θ^+ particles would decay close to the target and would not contribute to the K^+ beam examined in the usual emulsion exposures or in the K^+ lifetime experiments. With an incident beam consisting mostly of τ^+ particles, the lifetime measured for the $K^+_{\pi 2}$ mode would necessarily be closely equal to the lifetime for the 3π decay, since the $K^+_{\pi 2}$ decays observed would then be due to the secondary θ^+ particles. Similarly, the mass measurements obtained by the momentum–range method would be appropriate to the incident τ^+ particle, irrespective of the decay mode observed, so that the $K^+_{\pi 2}$ and τ^+ masses obtained in this way would necessarily be equal. On the other hand, the K particle masses obtained from the energy of the decay products would always give the mass of the immediate parent particle, i.e. the τ mass for the 3π decay and the θ mass for the $K_{\pi 2}$ decay. The two questions of immediate importance are then (*a*) how large a mass difference would be needed to give a radiative lifetime of order 10^{-8} sec as required from the measured τ^+ lifetime and the relative frequency of the $K^+_{\pi 2}$ decay, and whether this could be detected by comparison of the mass values obtained in these two ways, and (*b*) whether these γ rays could be directly observed. These are again the important questions if either (7.1*b*) or (7.2*b*) hold, with the θ^+ particle as the parent, the τ^+ as the short-lived daughter.

With zero spin for both τ and θ, the 2γ radiative decay is an allowed process, slowed down by a factor $(1/137)^2$ due to the two electromagnetic interactions and by a factor of the order $(kR)^4$ reflecting the weakness of the electromagnetic fields of the photons at the τ meson radius R for photons of long wavelength. The lifetime for the 2γ transition may be estimated as (Dalitz 1956 b)

$$T(\tau^+ \rightarrow \theta^+ + 2\gamma, \quad \text{or} \quad \theta^+ \rightarrow \tau^+ + 2\gamma) \simeq 1{\cdot}3 \times 10^{-3}/\eta^2 (m_\pi R)^4 \Delta^7 \text{ sec} \qquad \text{......(7.3)}$$

where η is a numerical constant of order unity and Δ is the mass difference (in electron masses). With the largest value of R reasonable (the pion Compton wavelength), a lifetime of 10^{-8} sec would then require a τ^+–θ^+ mass difference of at least $10 m_e$. From the mass measurements obtained from the observed energy release (cf. table 2), this possibility is clearly excluded.

If either τ or θ has zero spin, the other having spin 2, the transition (7.1) would be possible as a magnetic quadrupole transition. Treiman and Wyld (1955) have given an estimate of the corresponding lifetime, which shows that a mass difference of one or two MEV would be necessary. However, Alvarez *et al.* (1956 b) have looked for a γ ray of greater than 0·5 MEV emitted in coincidence with $K^+_{\pi 2}$ decay, after some minor modifications in the equipment used for their measurement of the $K_{\pi 2}$ lifetime. In fact a γ ray of about 0·5 MEV accompanied about 10% of

$K_{\pi 2}$ decays, but it was expected that annihilation radiation from electron showers due to the π^0 decays following $K^+_{\pi 2}$ decay should be counted with about this frequency. No evidence was found for a Lee–Orear γ ray corresponding to transitions (7.1*a*) or (7.2*a*). The possibility (7.1*b*) for which the γ ray would accompany the τ^+ decay has not yet been checked in this way.

Finally, if the spins were 2^+ and 2^-, an electric dipole transition (7.1) would be possible. The observed lifetime would require a mass difference of only several kev. This possibility is therefore not excluded by the present data, but there are no other reasons to favour this situation.

It has been pointed out by Weinstein (1956, unpublished) that the electric and magnetic fields effective on a K particle passing through matter could induce electromagnetic transitions between τ and θ states, provided the τ^+–θ^+ mass difference were extremely small (of order 10^{-5} ev or less). In this way a random mixture of τ and θ states would result by the time the particle came to rest, and the lifetime observed would be the lifetime of this mixture, which would naturally not depend on the decay mode observed. If the transition were electric dipole, this mixing could be achieved in a time of order 10^{-12} sec (Yang 1956) in condensed matter, taking a reasonable estimate for the dipole moment between τ and θ states. For spins 2^+ and 2^-, this could also provide an explanation for the lifetime identity. In this case, the lifetimes observed for decay in flight through vacuum and for decay after being brought to rest would generally differ, which would be reflected by variation of relative frequencies of various K^+ decay modes with time of flight to the emulsion; there is no evidence for this effect at present. However, if zero spin holds for either or both of τ and θ, then the coupling with the electric and magnetic fields in matter will be very weak, no appreciable mixing would occur in 10^{-8} sec, and Weinstein's suggestion would not contribute to the explanation for the τ^+–θ^+ lifetime identity.

7.2. *Non-conservation of Parity in Weak Decays?*

The argument that the τ^+ and θ^+ particles are distinct depended on the conclusion that the three pions from τ^+ decay are in a configuration with parity opposite from that appropriate to a K particle with a decay to two pions. This argument would not go through if conservation of parity were required only for the strong and electromagnetic interactions, and not for these weak decay interactions. In this case there could be a unique K^+ particle with many competing decay modes, the 2π and 3π modes leading to final pion states of opposite parity. This particle would then have a unique lifetime and all of the observed similarities amongst K^+ particles would have a very direct explanation.

The implications of the assumption that weak decay processes do not conserve parity have been considered by Lee and Yang (1956 b) and by Gatto (1957). Owing to the strong (parity-conserving) interactions of the K particle with Λ, Σ and Ξ particles, non-conservation of parity for the K particle decay would imply parity nonconservation also for the hyperon decay processes. Strongly interacting particles would then have definite parity only to a (very) good approximation, for these weak interactions would couple them to a small percentage of states of opposite parity; from the known strength of the decay interactions, the intensity of this admixture of opposite parity would be at most of the order of 10^{-12}. One result of such an admixture would be a non-zero electric dipole moment for the nucleons ;† however, the most accurate experiments to date (on the neutron's electric dipole moment (Purcell and Ramsey 1950)) would need improvement by a factor 10^6 to see an effect of the expected magnitude.

If parity conservation does not hold for these slow decay processes, then both s- and p-wave pions may result from Λ^0 (or Σ) decay, spin $\frac{1}{2}$ being assumed for the Λ^0 (or Σ) particle. These waves will generally interfere and give rise to characteristic features not permissible in an angular distribution resulting from a state of definite parity through parity-conserving interactions. Whereas, with parity conservation, the decay probability is a true scalar (being the square of a wave function of definite parity), here the interference terms between the final states of opposite parity which violation of parity conservation allow, will form a pseudoscalar term in the decay probability. Whereas, for Λ^0 spin $\frac{1}{2}$, an isotropic distribution of pions must result from a parity-conserving Λ^0 decay, the probability distribution for Λ^0 decay leading to final states of both parities can depend (linearly) on the pseudoscalar $\mathbf{P}.\mathbf{q}$, where $\mathbf{P}$ denotes the Λ^0 polarization vector and $\mathbf{q}$ the momentum vector of the outgoing pion. In the strong production process giving rise to the Λ^0 particle, parity is necessarily conserved, so that the Λ^0 polarization vector $\mathbf{P}$ must be normal to its production plane, this being the only pseudovector relevant to the production process. (Note : the polarization $\mathbf{P}$ is defined as the average of the spin vector in the beam, and spin, being an angular momentum, is a pseudovector.) Hence parity non-conservation for Λ^0 decay could lead to a backward–forward asymmetry (observed in the Λ^0 rest system) relative to the normal to the production plane for the Λ^0, even with Λ^0 spin $\frac{1}{2}$. No evidence for such an asymmetry has been found in the $(\pi^- + \mathrm{P})$ experiment of Budde, Chretien, Leitner, Samios, Schwartz and Steinberger (1956). However, the absence of an asymmetry could arise in many ways, even with non-conservation of parity in Λ^0 or Σ decay. The interference between states of opposite parity will be proportionately small if either s- or p-wave emission is dominant (e.g. p-wave emission may be relatively suppressed by the centrifugal barrier faced by the outgoing pion), or if the s- and p-wave amplitudes have a phase difference in the region of 90° ; alternatively the asymmetry could be small simply because the polarization $\mathbf{P}$ of the Λ^0 or Σ particles produced in the process considered may happen to be small.

With the possibility of parity non-conservation for strange-particle decays, it is natural to enquire whether other weak decays of comparable strength, such as the beta decays for neutron and muon, may not also involve a violation of parity conservation. Lee and Yang (1956 b) have pointed out that our experimental evidence on nuclear beta decay has not yet tested this point. The beta decay interaction terms which conserve parity and those which reverse parity are separately indistinguishable since one set may be obtained from the other simply by replacing neutrinos by antineutrinos (the neutrino and antineutrino have opposite parity), a change which we have no present means for detecting. Evidence for parity

† *Note added in proof* : However, Landau (1957, to be published) has pointed out that a non-zero electric dipole moment is excluded for nucleons or nuclei if time reversal invariance holds for the weak (as well as for the strong) interactions, irrespective of the parity non-conservation for weak interactions. This is readily seen from the fact that the electric dipole moment can only be parallel to the spin of the nucleus so that the interaction energy must have the form $D\mathbf{S}\cdot\mathbf{E}$ where $\mathbf{E}$ is the electric field at the nucleus. Under reversal of time, $\mathbf{S}$ is reversed whereas $\mathbf{E}$ remains unchanged so that this interaction is invariant under time reversal only if $D = 0$.

non-conservation could come only from interference between the two sets of interaction terms, involving again a pseudoscalar term in the decay probability. Such a pseudoscalar can occur most directly in the beta decay of an oriented nucleus, since the polarization vector $\mathbf{P}$ of the initial nucleus and the momentum $\mathbf{q}$ of the beta particle can form the pseudoscalar $\mathbf{P}.\mathbf{q}$. For an allowed beta transition from an oriented nucleus, the angular distribution of the beta particles will have the form $1+A(v/c)P\cos\theta$, where θ is the angle between $\mathbf{q}$ and the orientation direction $\mathbf{P}$, and v is the beta particle velocity. Neglecting axial coupling as small in beta decay (as experiment indicates), the numerical constant A is proportional to the interference product $(c_T{}^+c_T{}^-)$ between the interaction constants for the tensor interactions which respectively conserve $(c_T{}^+)$ and reverse $(c_T{}^-)$ total parity (Lee and Yang 1956 b). An experiment to measure the degree of forward–backward asymmetry in the beta decay of an oriented nucleus is at present in progress at the National Bureau of Standards. Its result should settle the question of parity conservation for nuclear beta decay, at least for the tensor interaction term.

Lee and Yang (1956 b) and Gatto (1957) have pointed out that the violation of parity conservation is only consistent with general principles (such as Lorentz invariance) if other symmetry principles are also violated for the weak interactions. This depends on a theorem first proved by Luders (1954), concerning which Pauli (1955) has recently given an illuminating discussion. What this theorem proves is that, for any Lorentz-invariant theory in which the normal connection between spin and statistics is assumed to hold for all particles, the invariance of the interactions under any two of the following operations, (i) particle–antiparticle conjugation (cf. § 5.3), (ii) (Wigner) time reversal,[†] (iii) spatial reflections (which implies parity conservation), requires that the interactions are necessarily invariant under the third operation. Consequently, the proposal to give up parity conservation for weak interactions (i.e. to give up invariance under spatial reflections (iii)) involves

---

[†] The operation of time-reversal introduced into quantum mechanics by Wigner (1932) is the direct analogue of the time-reversal of classical physics which corresponds simply to a relabelling of the time axis in the reverse direction. Wigner time-reversal of a given state involves reversal of the momentum, the angular momentum and the intrinsic spin for each particle, leaving unaffected its space coordinates, its charge, its isotopic spin, etc.; the wave function ψ^T for the time-reversed state is related to the initial wave function ψ by

$$\psi^T = C\psi^*, \qquad \ldots\ldots(i)$$

in which ψ^* denotes the complex conjugate of ψ and C is a unitary operator reversing the spin of each particle. If the Hamiltonian H for the system is invariant under Wigner time-reversal (formally this means $H^* = C^{-1}HC$), then

$$C\psi^*(m) = \lambda_m\psi(-m) \qquad \ldots\ldots(ii)$$

where m is the z-component of total angular momentum: the time-reversed state ψ^T necessarily corresponds to spin component opposite to ψ. In (ii), λ_m is a phase factor depending only on m; $\psi(m)$ is usually defined in such a way that $\lambda_m = (i)^{2m}$.

From (i) it is apparent that ψ^T is not related to ψ by a unitary transformation. Invariance of the interactions under Wigner time-reversal therefore does not lead to selection rules for processes but, with the relation (ii) then holding, to reality conditions for certain matrix elements in the theory. The most important of these is the condition that the elements of the K-matrix (the reactance matrix, or the transition matrix calculated assuming always standing waves at infinity) should be real: this follows irrespective of the validity of other symmetry operations such as particle–antiparticle interchange or space-reflections. This condition has been demonstrated most generally by Gell-Mann and Watson (1954). From this, the relative phases between certain matrix elements can be predicted from a knowledge of the interactions occurring between the initial and between the final particles. This result has been used in the discussion of strange particle decay modes by Takeda (1956), see §§5.1 and 6.1. If the initial system is in a stationary state and the interactions in the final state may be neglected, it is the transition amplitude itself which is real; this restriction has been used in the discussion of K_2^0 decay by Treiman and Sachs (1955; cf. §5.4) and has been discussed for nuclear beta decay by Pauli (1955) and by earlier authors.

giving up also invariance with respect to either particle–antiparticle conjugation or time reversal.

The assumption of invariance under time reversal has generally formed the basis from which the principle of detailed balancing has been shown to follow (Coester 1951) so that there is naturally considerable reluctance to give up this invariance property, although there is naturally no empirical evidence that detailed balancing must hold for these slow processes. Also time-reversal invariance in the quantum theory of particles is a direct generalization of the time reversal invariance which is known to hold in all classical theories. Invariance under time reversal leads to quite definite statements concerning the relative phases between matrix elements for certain weak decay processes, as has been remarked at various points above (e.g. §§ 5.1, 5.4 and 6.1). These statements have been of great value in simplifying the analysis of the empirical data but, at the present time, these data do not provide any check on the assumption of time-reversal invariance for weak decay processes.

On the other hand, Lee and Yang (1956 b) have pointed out that the observed equality of masses and lifetimes as between particles and antiparticles (for the $K^{\pm}, \pi^{\pm}, \mu^{\pm}$ in particular) provides no compelling argument for the invariance of weak decay processes under particle–antiparticle conjugation. Equality of lifetimes for particle and antiparticle can be shown to follow from the requirement that the interaction energy for the decay processes be real as well as invariant under proper Lorentz transformations, the strong interactions naturally being assumed invariant with respect to all three operations specified above. This result holds only to lowest order in the weak interactions, which is a more than adequate approximation for any practical case since higher approximations in the weak interactions involve coupling parameters of order 10^{-12}. Similarly, weak interactions which are not invariant with respect to particle–antiparticle conjugation can produce only a negligible mass difference ($\lesssim 10^{-5}$ ev) between particle and antiparticle. It therefore appears permissible to consider that parity conservation and invariance under particle–antiparticle conjugation may both fail for the weak decay interactions. With invariance under time reversal, the interaction energies will then be invariant with respect to the joint operation of space reflection and particle–antiparticle conjugation, according to Luders' theorem, although not invariant with respect to the separate operations.

It is of interest to note here that the θ_1^0, θ_2^0 situation discussed by Gell-Mann and Pais (1955) (cf. §§ 5.2 and 5.3) may still hold, at least to a fairly good approximation, even when invariance under space reflections and under particle–antiparticle conjugation is no longer the case. This is due to the fact that the $\theta^0 \rightarrow 2\pi$ coupling (5.1) is by far the strongest of the decay interactions, and that two pions with given total angular momentum J necessarily have a definite space parity ($w = (-1)^J$) and charge conjugation parity ($c = (-1)^J$). With $J = 0$, invariance under time reversal requires that the matrix elements for the interactions ($\theta^0 \rightarrow 2\pi$) and ($\bar{\theta}^0 \rightarrow 2\pi$) be equal, from which it follows as before that only the θ_1^0 state can undergo decay to two pions, the θ_2^0 state being long-lived in comparison.

Parity non-conservation corresponds to a left–right asymmetry in our world. Although this possibility is intuitively distasteful, there seems no *a priori* reason why such an asymmetry should not exist for weak decay processes ; in an anti-world, consisting of antiprotons, antineutrons, etc., an opposite left–right asymmetry would hold for weak decay processes, according to the theorem of Luders (invariance under time reversal being assumed). This possibility would allow the most direct resolution of the τ–θ problem, with only the one (very far-reaching) assumption; however, empirical evidence explicitly demonstrating the failure of parity conservation in hyperon or K particle decays will be necessary before this hypothesis could be accepted with conviction.

7.3. *The Hypothesis of Invariance under Parity Conjugation*

As emphasized by Lee and Yang (1956 a), the degree of equality observed for τ^+ and θ^+ masses (these being assumed distinct particles) strongly suggests the existence of some symmetry principle effective for all strong interactions. Otherwise, this mass degeneracy could only be viewed as an extremely improbable coincidence. As a result of their parity difference, the τ and θ particles would in general have strong interactions of quite different forms; these interactions would produce large and unrelated contributions to the masses of τ and θ, relative to which a mass difference of less than one MEV is quite negligible.

The simplest symmetry principle which ensures this mass equality consists of the requirement that all strong interactions remain unchanged when τ and θ particles are interchanged. Clearly this principle is possible only if the τ and θ particles have the same spin value, as we will now assume. Lee and Yang (1956 a) refer to this operation of interchanging τ and θ as 'parity conjugation' and denote it by the symbol C_p. It is natural to suppose that pions and nucleons are left unchanged by this operation C_p.

Consider the effect of this 'parity-conjugation' operation on a strong interaction involving a θ^+ particle, for example

$$\pi^+ + \mathrm{N} \rightarrow \Lambda^0 + \theta^+. \qquad \ldots\ldots(7.4a)$$

This symmetry principle now requires that there exists a reaction with equal amplitude, obtained by the parity conjugation of each particle of (7.4*a*). This parity-conjugate reaction will thus be

$$\pi^+ + \mathrm{N} \rightarrow \Lambda_2{}^0 + \tau^+, \qquad \ldots\ldots(7.5)$$

where $\Lambda_2{}^0$ is the particle which is parity conjugate to the Λ^0 particle of (7.4*a*), which we will henceforth denote by $\Lambda_1{}^0$. Since the spin and orbital states of the various particles are not affected by this operation, the $\Lambda_2{}^0$ particle of (7.5) must necessarily have intrinsic parity opposite to that for the $\Lambda_1{}^0$ particle in order that the parity of the final state be unchanged by C_p. The existence of a Λ^0 doublet whose two members have the same spin, isotopic spin and strangeness, but opposite intrinsic parity, is therefore a direct consequence of this symmetry principle. The reaction (7.4*a*) will now be written

$$\pi^+ + \mathrm{N} \rightarrow \Lambda_1{}^0 + \theta^+. \qquad \ldots\ldots(7.4b)$$

It is also permitted by all of the present conservation laws that the $\Lambda_2{}^0$ particle may be produced together with the θ^+ particle in a $(\pi^+ + \mathrm{N})$ collision, thus

$$\pi^+ + \mathrm{N} \rightarrow \Lambda_2{}^0 + \theta^+, \qquad \ldots\ldots(7.6)$$

but the symmetry principle makes no statement concerning the relative strength of these reactions (7.4*b*) and (7.6). To reaction (7.6), there will necessarily exist a parity-conjugate reaction with equal amplitude, namely,

$$\pi^+ + \mathrm{N} \rightarrow \Lambda_1{}^0 + \tau^+. \qquad \ldots\ldots(7.7)$$

On the same basis, it will be apparent that this principle requires a 'parity doublet' also for the Σ particles, which will therefore appear in both the Σ_1 and Σ_2 varieties. On the other hand, the Ξ particle can be produced together with two K particles from a pion–nucleon system. Under the operation C_p the parity of the final state is preserved if the particle which is parity-conjugate to the Ξ particle has the same parity as the Ξ particle, as well as the same spin, isotopic spin, strangeness, etc. It is therefore natural to identify the particle parity-conjugate to the Ξ particle with the Ξ particle itself, so that this will be an eigenstate of the operation C_p. However, there are still two possibilities, $C_p\Xi = \pm\Xi$, and the sign appropriate to the physical case may be deduced, at least in principle, from experimental observations on $(\Xi + \mathrm{K} + \mathrm{K})$ production processes.

Thus for every reaction involving strange particles, there will be a reaction of equal amplitude relating the parity-conjugate particles. In particular, if the initial system is unchanged by the operation C_p, then for every reaction leading to a final state containing a certain strange particle there will necessarily be a reaction of equal amplitude leading to the parity conjugate final state containing the particle parity-conjugate to this strange particle. For a pion–nucleon or nucleon–nucleon collision, therefore, the production cross section for a given strange particle (summed over all possibilities for the other particles produced) must be equal to the production cross section for the parity-conjugate particle. For example, it follows that in such collisions the τ/θ ratio at production is necessarily unity, whence τ and θ production necessarily follow the same excitation function. This may be illustrated by the reactions given above, which are the only strange-particle reactions for a $(\pi^+ + \mathrm{N})$ collision sufficiently near threshold. The intensity of θ^+ production through reactions (7.4*b*) and (7.6) together is equal to the intensity of τ^+ production through (7.5) and (7.7), since the reaction amplitudes for (7.4*b*) and (7.5) are equal, as are also those for (7.6) and (7.7). A similar situation holds for $\Lambda_1{}^0$ and $\Lambda_2{}^0$ production provided the nature of the accompanying K particle is not taken into account, since amplitudes (7.4*b*) and (7.7) leading to $\Lambda_1{}^0$ particles are equal to those for (7.5) and (7.6) leading to $\Lambda_2{}^0$ particles. On the other hand, if account is taken of the nature of the accompanying K particle, this equality of production no longer holds. For example, the intensity of $\Lambda_1{}^0$ production together with θ^+ particle will generally not equal the intensity of $\Lambda_2{}^0$ production with accompanying θ^+ particle. Since the spin and angular momentum states effective in these two processes are necessarily different (the intrinsic parity being reversed for only one of the final particles), these processes are not related by the symmetry principle; indeed, this remark indicates that the ratio between these amplitudes must generally be energy dependent, especially near threshold where the relative effectiveness of different orbital states is most marked.

A further example is provided by the nuclear scattering of τ^+ and θ^+ particles. Scattering may occur with or without a change in intrinsic parity for the

K particle, thus

$$\tau^+ + \mathrm{P} \begin{array}{ll} \nearrow \tau^+ + \mathrm{P}, & \ldots\ldots(7.8a) \\ \searrow \theta^+ + \mathrm{P}, & \ldots\ldots(7.8b) \end{array}$$

$$\theta^+ + \mathrm{P} \begin{array}{ll} \nearrow \theta^+ + \mathrm{P}, & \ldots\ldots(7.9a) \\ \searrow \tau^+ + \mathrm{P}, & \ldots\ldots(7.9b) \end{array}$$

the processes (7.8*a*), (7.9*a*) and (7.8*b*), (7.9*b*) forming parity-conjugate pairs whose reaction amplitudes are equal. The differential cross sections for τ^+ and θ^+ scattering are therefore equal, and this will be true generally for nuclear scattering of τ and θ particles since the two initial systems form a parity-conjugate pair.

If the τ/θ ratio is unity for the incident K^+ beam, the τ/θ ratio will then continue to have the value unity for any scattered K^+ beam resulting from nuclear collisions. However, if the τ/θ ratio for the incident beam differs from unity, the τ/θ ratio will generally be modified by nuclear scattering to a degree depending on the ratio of amplitudes (7.8*a*) and (7.8*b*). These amplitudes will have an energy-dependent ratio since the change from τ^+ to θ^+ in reaction (7.8*b*) requires a change of one unit in the orbital angular momentum of the K^+ particle, whereas orbital angular momentum is unchanged for (7.8*a*). For low incident energy, the τ–θ process (7.8*b*) may be expected to have small probability relative to the direct scattering (7.8*a*), since the latter can occur entirely in the s-wave whereas an incident s-wave τ^+ particle leads to an outgoing θ^+ particle in the p-wave, which is strongly suppressed at low energy owing to the centrifugal barrier the outgoing θ^+ particle has to penetrate. On this basis the observation that the relative frequencies of various K^+ decay modes are essentially unaffected by nuclear scattering of the K^+ beam could reflect either (i) that the τ/θ ratio in the incident beam is close to unity, or (ii) that the scattering amplitudes for change from τ to θ (and vice versa) are small relative to the direct scattering amplitudes. The situation (i) would be in accord with a τ/θ ratio of unity at production, since the observed lifetimes for $\mathrm{K}^+_{\pi 2}$ and τ^+ decay allow the τ/θ ratio to change only slightly in the flight time relevant to the emulsion observations.

The parity-conjugation symmetry principle naturally provides no explanation of the lifetime equality observed for τ^+ and θ^+ decay, since it relates only to the strong interactions: in any case, τ^+ and θ^+ decays lead to states containing different numbers of pions, which are obviously not related through any such symmetry principle. At the moment this equality can be considered only as an approximate coincidence, from this viewpoint; the experimental data still allows up to about 20% difference between τ^+ and $\mathrm{K}^+_{\pi 2}$ lifetimes. Bludman (1956) has remarked that this lifetime equality might appear less coincidental if the predominant K^+ decay mode, the $\mathrm{K}^+_{\mu 2}$ mode (to which is added the $\mathrm{K}^+_{\mu 3}$ mode), could result from the decay of either τ or θ particle. This would be possible if, for example, the τ^+ particle had the decay mode

$$\tau^+ \to \mu^+ + \bar{\nu}. \qquad \ldots\ldots(7.10a)$$

analogous to the π–μ decay, whilst the θ^+ particle had a decay

$$\theta^+ \to \mu^+ + \nu, \qquad \ldots\ldots(7.10b)$$

where $\bar{\nu}$ denotes an antineutrino, which has necessarily an intrinsic parity opposite to that of the neutrino. Bludman pointed out further that these decay processes

(7.10*a*, 7.10*b*) could have the same amplitude and therefore equal partial lifetimes, and that this would not be an unreasonable situation. If the τ/θ ratio is unity at production and the τ^+ and θ^+ lifetimes are assumed identical, the K^+ beam will consist equally of τ^+ and θ^+ particles; Bludman's hypothesis would then imply that the $K^+_{\mu 2}$ (and $K^+_{\mu 3}$) events observed in emulsion exposures arise equally from the decay of τ^+ and θ^+ particles. Since one-half of the K^+ decays are to be attributed to the decay of θ^+ particles, it follows that the number of θ^+ particle decays leading to other than the $K^+_{\mu 2}$ and $K^+_{\mu 3}$ modes should equal the number of τ^+ particle decays which are not of the $K^+_{\mu 2}$ or $K^+_{\mu 3}$ type. In actual fact the proportion of K^+ decays leading to the $K^+_{\pi 2}$ and to the $\tau + \tau'$ modes are about 29% and 8% respectively. The discrepancy between these numbers may be reduced slightly by attributing all K^+_{e3} events to the τ^+ particle (Schwinger 1956). Although there is still some doubt concerning the relative frequency of K^+_{e3} decay (values between 2% and 9% have been given by various authors (cf. table 4)), it seems unlikely that this could be sufficiently high ($\sim 15\%$) to bring the ratio $K^+_{\pi 2}/(\tau + \tau' + K^+_{e3})$ up to unity, as required for equal lifetimes and the Bludman hypothesis. If the relative frequencies given by Birge, Perkins, Peterson, Stork and Whitehead (1956) are accepted, this situation is not appreciably improved by allowing any values for τ^+ and θ^+ lifetimes compatible with the experimental values of Fitch and Motley (1956); these relative frequencies would be compatible with the Bludman hypothesis only if the τ^+ lifetime were about three times shorter than the θ^+ lifetime, which is certainly excluded. This particular interpretation of the τ–θ lifetime equality appears far from hopeful at the present time. This equality thus remains before us as an accidental coincidence, for which no explanation is yet available. In this situation, the experimental relationship between $\tau_1{}^0$ and $\theta_1{}^0$ lifetimes and between $\tau_2{}^0$ and $\theta_2{}^0$ lifetimes will be of the greatest interest, since an equality between these lifetimes could scarcely also be attributed to an independent accidental coincidence.

Different lifetimes are also expected for the decay of $\Lambda_1{}^0$ and $\Lambda_2{}^0$ particles. Both of these particles will undergo a visible decay to a $(\pi^- + P)$ system with an energy release of 37 MEV, so that it cannot be decided whether an individual Λ^0 event represents $\Lambda_1{}^0$ or $\Lambda_2{}^0$ decay. However, the Λ^0 particle of positive parity, which we shall denote by $\Lambda_+{}^0$, leads to a p-wave pion in its decay, whereas decay of the Λ^0 particle of negative parity $(\Lambda_-{}^0)$ leads to an s-wave pion (assuming spin $\frac{1}{2}$ for the Λ^0 particles). Since the outgoing pion in $\Lambda_+{}^0$ decay faces a centrifugal barrier, it is natural to expect the $\Lambda_+{}^0$ lifetime to be longer than that of the $\Lambda_-{}^0$ decay. The observed Λ^0 lifetime (cf. § 6.1) may then be attributed to the $\Lambda_-{}^0$ decay; since these lifetime measurements have all been made close to the origin of the Λ^0 particles, the absence of evidence for a Λ^0 particle of longer lifetime is not yet significant, since the long-lived Λ^0 particle would escape from the sensitive region before decay.

A similar situation should hold also for the $\Sigma_+{}^\pm$ and $\Sigma_-{}^\pm$ particles, with different lifetimes for even and odd parity particles. However, this effect would be immediately apparent in the experiments on K^- capture in hydrogen if these lifetimes were widely different, for the long-lived Σ^- particles in particular may be clearly identified even though they escape from the chamber before decay. The experimental decay curve for Σ^- particles resulting from K^-–proton capture (Solmitz, Alvarez, Bradner, Falk, Gow, Rosenfeld and Tripp 1956) shows no indication for the existence of two lifetimes, although it could not exclude two lifetimes differing by as much as a factor 3.

The nuclear interactions of Λ_+^0 and Λ_-^0 particles must be identical, according to the parity-conjugation symmetry principle. The Λ hypernuclei will therefore appear also as parity doublets and the two species will appear with equal frequency, at least from pion or nucleon induced stars or from K^- capture stars if the incident K^- beam has a τ/θ ratio of unity. Now, in addition to an interaction between $\Lambda_\pm^0$ particle and nucleon in which the intrinsic parity of the Λ^0 is unchanged, thus

$$\Lambda_\pm^0 + \mathrm{P(N)} \rightarrow \mathrm{P(N)} + \Lambda_\pm^0, \qquad \ldots\ldots(7.11a)$$

there is also a possibility of an interaction in which the intrinsic parity of the Λ^0 particle is reversed,

$$\Lambda_\pm^0 + \mathrm{P(N)} \rightarrow \mathrm{P(N)} + \Lambda_\mp^0. \qquad \ldots\ldots(7.11b)$$

However, if the initial $\Lambda_\pm^0$ particle is in an s-state relative to the nucleon, the final $\Lambda_\mp^0$ particle in (7.11*b*) must be in a p-state. Treiman (1956 b) has pointed out that the ground states of hypernuclei would then be a superposition of a state with a $\Lambda_\pm^0$ particle in an s-state relative to the nuclear well together with a state representing a $\Lambda_\mp^0$ particle in a p-state in this well. If the admixture of the latter state is appreciable, then the lifetimes observed for even the lightest hypernuclei ${}_{\Lambda_+}H^3$ and ${}_{\Lambda_-}H^3$ would agree with neither the Λ_+^0 nor the Λ_-^0 lifetimes, but would each lie intermediate between these values, since the hyperon in these hypernuclei would be a Λ_+^0 hyperon for a certain fraction of time and a Λ_-^0 hyperon for the remainder. Indeed, a reasonable estimate for the degree of admixture could be obtained from lifetime measurements for ${}_\Lambda H^3$ hypernuclei.

The existence of two varieties of Λ hypernuclei also requires a re-examination of the Ruderman–Karplus argument relating the internal conversion coefficient for non-mesonic decay of hypernuclei with the Λ^0 spin–parity. Firstly, it must be taken into account that the experimental observations measure only the average internal conversion for Λ_+^0 and Λ_-^0 hypernuclei. Since the Λ_+^0 and Λ_-^0 hypernuclei necessarily appear with equal frequency, this average coefficient R is related to the separate coefficients R_+ and R_- by the expression

$$R = (2R_+R_- + R_+ + R_-)/(R_+ + R_- + 2). \qquad \ldots\ldots(7.12)$$

If either R_+ and R_- considerably exceeds unity, then the observed value of R will be about twice the lesser of R_+ and R_-. If the estimates of R_l given by Ruderman and Karplus are accepted, then the data would be consistent with spin $\frac{1}{2}$ for the Λ^0 parity doublet since the expected value of R would be about twice the value R_0 given in § 6.4, which is still compatible with the empirical data. A second factor also modifying the Ruderman–Karplus argument has been pointed out by Treiman (1956 b). In the calculation of the internal conversion coefficient R_l for a $\Lambda_\pm^0$ particle of given spin and parity, the admixture of p-wave state in the $\Lambda_\pm^0$ hypernucleus must be taken into consideration since its contribution to the internal conversion process will be quite different from that of the s-wave $\Lambda_\pm^0$ state. However, it appears unlikely that this effect would destroy the agreement found with spin $\frac{1}{2}$ for the Λ^0 particle, or allow agreement with any higher spin value.

The decay processes for parity-conjugate hypernuclei will necessarily be different since the final systems necessarily have opposite parity. A particularly important illustration of this is provided by the case of the ${}_{\Lambda+}\mathrm{H}^4$ and ${}_{\Lambda-}\mathrm{H}^4$ hypernuclei. Since these hypernuclei have the same spin but opposite parity, it follows that only one of these hypernuclei can undergo the $(\pi^- + {}^4\mathrm{He})$ decay process by which the ${}_{\Lambda}\mathrm{H}^4$ hypernucleus was first established, since only one parity is possible for this final system for given total angular momentum. For example, if antiparallel spins are favoured in the Λ^0–nucleon interaction, then the observed ${}_{\Lambda}\mathrm{H}^4$ hypernucleus will be the ${}_{\Lambda-}\mathrm{H}^4$ hypernucleus of spin $0-$, from which the $(\pi^- + {}^4\mathrm{He})$ decay can take place. This still allows the existence of a ${}_{\Lambda+}\mathrm{H}^4$ hypernucleus of spin $0+$ for which only the other decay modes such as $(\pi^- + \mathrm{P} + {}^3\mathrm{H})$ can occur.

The assumption of a $\frac{1}{2} \pm \Lambda^0$ parity doublet therefore does not necessarily conflict with the data on Λ hypernuclei. The Ruderman–Karplus argument indicates the least of the l values effective in $\Lambda_+{}^0$ and $\Lambda_-{}^0$ decay, which is then $l = 0$ corresponding to spin $\frac{1}{2}-$. On the other hand, the binding energy data, following the arguments of §6.2, indicate the spin–parity of the Λ^0 particle which can form a ${}_{\Lambda}\mathrm{H}^4$ hypernucleus capable of $(\pi^- + {}^4\mathrm{He})$ decay; this may be $\frac{1}{2}+$ if the ${}_{\Lambda}\mathrm{H}^4$ spin is shown to be $1+$, or $\frac{1}{2}-$ if this hypernucleus has spin $0-$.

Lee and Yang (1956 c) and Gatto (1956 c) have pointed out some interference phenomena whose observation would indicate the existence of a Λ (or Σ) parity doublet. These effects will be illustrated by considering the $(\pi^+ + \mathrm{N})$ reactions leading to a θ^+ particle, as given in equations (7.4*b*) and (7.6). For a given direction of Λ^0 motion, the $\Lambda_+{}^0$ and $\Lambda_-{}^0$ states produced will have a definite phase relation, thus

$$\psi_\theta(\Lambda^0, t = 0) = a\psi_\theta(\Lambda_+{}^0, t = 0) + b\psi_\theta(\Lambda_-{}^0, t = 0), \qquad \ldots\ldots(7.13)$$

since the accompanying θ^0 particle (assumed spin 0^+) is in the same state for either case. After a flight time t, the state (7.13) will develop to

$$\begin{aligned}\psi_\theta(\Lambda^0, t) = [&a \exp(-\lambda_+ t/2)\, \psi_\theta(\Lambda_+{}^0, 0) \\ &+ b \exp(-\lambda_- t/2) \exp(it\Delta)\, \psi_\theta(\Lambda_-{}^0, 0)] \exp(i\omega_+ t) \qquad \ldots\ldots(7.14)\end{aligned}$$

where λ_+ and λ_- are the decay constants for $\Lambda_+{}^0$ and $\Lambda_-{}^0$ particles, ω_+ is the mass energy of the $\Lambda_+{}^0$ particle and Δ is essentially $m_- - m_+$, the mass difference between $\Lambda_+{}^0$ and $\Lambda_-{}^0$. If S_+ and S_- now denote the amplitudes for $\Lambda_+{}^0$ and $\Lambda_-{}^0$ decay to a p-wave and an s-wave π^- meson respectively (spin $\frac{1}{2}$ is assumed here for the Λ^0 doublet), the final $(\pi^- + \mathrm{P})$ system (specified in the Λ^0 rest frame) occurring at the point reached after flight time t is

$$\begin{aligned}\psi_\theta(\pi^- + \mathrm{P}) = [&a S_+ \exp(-\lambda_+ t/2) |\pi^- + \mathrm{P}\rangle_{\mathrm{P}} \\ &+ b S_- \exp(-\lambda_- t/2) \exp(it\Delta) |\pi^- + \mathrm{P}\rangle_{\mathrm{S}}]. \qquad \ldots\ldots(7.15)\end{aligned}$$

These two terms will generally interfere in the probability for emission of a π^- meson in a given direction. The π^- angular distribution will have the form

$$\begin{aligned}F_\theta(\cos\chi) = [&a^2 S_+{}^2 \exp(-\lambda_+ t) + b^2 S_-{}^2 \exp(-\lambda_- t) \\ &+ 2\cos\chi\, \mathscr{R}\{a S_+ b^* S_-{}^* \exp(it\Delta)\}] \exp[-(\lambda_+ + \lambda_-)t/2], \\ &\qquad \ldots\ldots(7.16)\end{aligned}$$

where χ is the angle between the π^- meson and the Λ^0 direction of motion.

The Λ^0 wave function $\psi_\tau(\Lambda^0, t = 0)$ resulting from τ^+ production in this $(\pi^+ + \mathrm{N})$ collision may be obtained by interchanging *a* and *b* in (7.13), according to the parity-conjugation symmetry principle. The angular distribution $F_\tau(\cos\chi)$ is then obtained from (7.16) by interchanging *a* and *b*. The distribution of Λ^0 decays accompanied by either τ^+ or θ^+ production is simply given by the sum $[F_\theta(\cos\chi)+F_\tau(\cos\chi)]$.

The expression (7.16) predicts two types of effect arising from the existence of the Λ^0 parity doublet:

(*a*) The angular distribution of Λ^0 decays (observed in the Λ^0 rest system) occurring after a given time of flight will not generally be symmetrical forwards and backwards relative to the direction of motion of the Λ^0 particles, owing to the occurrence of interference between the s- and p-wave pions from $\Lambda_-{}^0$ and $\Lambda_+{}^0$ decay. This asymmetry can occur even though all the interactions effective are parity-conserving, because the final state can be reached from the initial state through alternative intermediate states whose parities are opposite.

(*b*) For a particular π^- angle of emission, the distribution of the decays with respect to flight time will not be a simple sum of two exponential curves corresponding to $\Lambda_+{}^0$ and $\Lambda_-{}^0$ decays separately, but will require an additional term with time dependence $\exp[(-\frac{1}{2}(\lambda_+ +\lambda_-)t]$. However, this additional term will vanish if all directions of π^- emission are included.

However, these effects may be small or difficult to observe for a variety of reasons. For example, if the mass-difference Δ is large relative to λ_-, the decay constant of the shorter-lived Λ^0 particle, the factor $e^{it\Delta}$ will cause the coefficient of $\cos\chi$ to oscillate rapidly with flight time—observation of the effect would then require very detailed observations, especially as they would be smeared out unless the Λ^0 momenta were accurately defined (by an accurately defined incident π^- energy).

Similar interference effects would be expected in the decay of Σ particles. These have been sought especially in the data on $\Sigma^\pm$ decay following K^- capture in hydrogen. The observed angular distributions have been given in § 4.7 and they are quite compatible with forward–backward symmetry. However, as just remarked above, the absence of these interference effects could arise in many ways even with the existence of a Σ parity doublet.

To sum up, the hypothesis of parity-conjugation symmetry appears necessary to account for the τ–θ mass equality, if these are assumed to be distinct particles (as is required if parity conservation holds for weak decays). This involves the existence of a similar multiplicity of Λ and Σ hyperons. This hyperon multiplicity is not yet inconsistent with any data at present available and would even allow a direct interpretation for the observations on Λ hypernuclei; on the other hand, there is at present no direct and clear empirical evidence which requires the existence of hyperon parity doublets.

7.4. *Schwinger's Model for Weak Decays*

In § 9 recent proposals by Schwinger (1956) concerning the strong interactions between pions, nucleons, K particles and hyperons will be mentioned briefly. One of the interactions introduced is a direct π–K coupling (cf. discussion in § 4.5)

$$\tau \longleftrightarrow \theta+\pi \qquad \ldots\ldots(7.17)$$

with coupling parameter $g_{\pi K}/m_\pi$, for which some empirical justification has been emphasized by Schwinger (cf. § 3.1). For our present purpose the important point is that the other strong interactions chosen by Schwinger are such that they cannot themselves give rise to a π–K interaction through any sequence of virtual processes; the complete π–K interaction is concentrated in this one term (7.17).

To provide a model for the weak decay processes leading to pions and nucleons, Schwinger introduces a weak coupling (with parameter λ) by means of which a certain K particle may simply disappear. Clearly this K particle must have zero charge, zero spin, even parity and charge-conjugation parity $+1$; all of these quantum numbers are required to be conserved rigorously and these are their values for the vacuum left after the disappearance of this K particle. Only the $\theta_1{}^0$ particle has these quantum numbers, so that the basic decay coupling proposed is

$$\theta_1{}^0 \xrightarrow[\lambda]{} \text{vacuum}. \qquad \ldots\ldots(7.18)$$

As a result of the π–K interaction, a K particle will have a pion cloud surrounding its core. When a $\theta_1{}^0$ particle in this core disappears, following (7.18), pions from this cloud are available to conserve energy, momentum, angular momentum and parity in the process and to produce the observed decay configurations. In this way the observed $\theta_1{}^0$ decay may be pictured as due to the following sequence of virtual processes

$$\theta_1{}^0 \xrightarrow[g_{\pi K}]{} \tau^+ + \pi^- \xrightarrow[g_{\pi K}]{} \theta_1{}^0 + \pi^+ + \pi^- \xrightarrow[\lambda]{} \pi^+ + \pi^-. \qquad \ldots\ldots(7.19)$$

Any decay process such as (7.19) which occurs in consequence of (7.18) will necessarily involve a change of strangeness $\Delta s = \pm 1$ in the system, since the $\theta_1{}^0$ particle is a superposition of states with $s = +1$ and $s = -1$, whereas the vacuum has $s = 0$. The rule (2.28) empirically observed is therefore guaranteed in this model. Similarly, from (7.18), the disappearance of a $\theta_1{}^0$ particle leads to a change $\Delta T = \frac{1}{2}$ in the total isotopic spin of the system. As discussed in §§ 4.4 and 5.1, this $\Delta T = \frac{1}{2}$ rule for weak decays, first proposed by Gell-Mann (1955), leads to a simple explanation for the striking difference between $\theta_1{}^0$ and θ^+ lifetimes and for the observed τ'/τ ratio. Of course, these rules could be introduced in many other ways, but Schwinger's proposal has the advantage of being direct and unambiguous.

Decay of the τ^+ particle may be pictured in the following sequence

$$\begin{aligned}\tau^+ &\xrightarrow[g_{\pi K}]{} \theta^0 + \pi^+ \xrightarrow[g_{\pi K}]{} \tau^+ + \pi^+ + \pi^- \\ &\xrightarrow[g_{\pi K}]{} \theta_1{}^0 + \pi^+ + \pi^- + \pi^+ \xrightarrow[\lambda]{} \pi^+ + \pi^- + \pi^+. \qquad \ldots\ldots(7.20)\end{aligned}$$

For a pseudoscalar τ^+ particle, the matrix element for the 3π decay (7.20) obtained from this model has quite considerable dependence on the negative pion energy.

If the π–K interaction is only moderately strong, the perturbation approximation will be reasonably good for this matrix element $S(\tau^+)$ for τ^+ decay; this gives essentially

$$\begin{aligned}\frac{S(\tau^+)}{S(\theta_1{}^0)} &= \frac{\sqrt{2}g_{\pi K}}{2M_\tau}\left(\frac{1}{\omega_1}+\frac{1}{\omega_2}\right) \\ &= \frac{\sqrt{2}g_{\pi K}}{2M_\tau}\,\frac{(M_\tau-\omega_{\rm p})\,\omega_{\rm p}}{\omega_1\,\omega_2\,\omega_{\rm p}}, \qquad \ldots\ldots(7.21)\end{aligned}$$

using the notation of § 4.3. The dominant energy dependence is provided by the factor $\omega_{\rm p}(M_\tau - \omega_{\rm p})$ which increases with increasing π^- energy. This gives a quite adequate fit to the energy dependence found in the empirical data, as is shown in figure 10. The known ratio between the partial lifetime for τ^+ decay and the

$\theta_1{}^0$ lifetime leads to an estimate $g^2_{\pi K}/m^2_\pi \sim 1/6$ for the π–K coupling parameter. The small value obtained suggests that the perturbation treatment assumed in (7.21) may have some degree of validity.

For $K^+_{\mu 2}$ decay, Schwinger proposes a direct interaction $\theta^+ \rightarrow \mu^+ + \bar{\nu}$. With the π–K coupling (7.17), the $K^+_{\mu 3}$ decay would then result from the sequence

$$\tau^+ \rightarrow \theta^+ + \pi^0 \rightarrow \mu^+ + \bar{\nu} + \pi^0.$$

Alternatively, an interaction $\tau^+ \rightarrow \mu^+ + \nu$ could exist, with a consequent $K^+_{\mu 3}$ decay

$$\theta^+ \rightarrow \tau^+ + \pi^0 \rightarrow \mu^+ + \nu + \pi^0.$$

In either case, the relationship of $K^+_{\mu 3}$ and $K^+_{\mu 2}$ decays would be similar to the relation between the τ^+ and $\theta_1{}^0$ decays above, the ratio $S(K^+_{\mu 3})/S(K^+_{\mu 2})$ being $g_{\pi K}/2M_\tau \omega_\pi$. Although phase space for $K_{\mu 3}$ decay is a little larger than for $\tau \rightarrow 3\pi$ decay, this ratio of matrix elements is rather smaller than (7.21), and corresponds to a relative frequency $K^+_{\mu 3}/K^+_{\mu 2} \sim 1/600$, in contrast with the observed ratio 1/20. This discrepancy may mean that the π–K coupling is really very strong and that other effects play a role in the $\tau \rightarrow 3\pi$ decay. Another possibility is that the $K_{\mu 2}$ and $K_{\mu 3}$ decay modes may be the indirect result of a Fermi interaction $\Lambda^0 \rightarrow P + \mu^- + \nu$ similar to the beta-decay interaction, with $K^+_{\mu 2}$ decay resulting from the sequence

$$\theta^+ \rightarrow \overline{\Lambda^0} + P \rightarrow \mu^+ + \bar{\nu}. \qquad \ldots\ldots(7.22)$$

This last possibility has been discussed particularly by Costa and Dallaporta (1955) and by Gell-Mann (1956).

For the K^+_{e3} decay mode, Schwinger does consider a mechanism similar to (7.22), based on the hypothesis of a beta interaction for the Λ^0 particle. One of the many possibilities has some special interest; this depends on the assumption that only the Λ^0 particle which has a strong direct coupling with the τ particle possesses a beta interaction

$$\Lambda^0 \rightarrow P + e^- + \nu \qquad \ldots\ldots(7.23)$$

and that this interaction involves only scalar and tensor combinations for the $e^- + \nu$ system, as is known for nuclear beta decay. Then the decay $\tau^+ \rightarrow e^+ + \bar{\nu}$ is forbidden by parity conservation since this would require a pseudoscalar $e^+ + \bar{\nu}$ system and none is available according to the preceding assumption. Hence (7.22) can lead only to the K^+_{e3} decay

$$\tau^+ \rightarrow \overline{\Lambda^0} + P \rightarrow \overline{\Lambda^0} + P + \pi^0 \rightarrow e^+ + \bar{\nu} + \pi^0. \qquad \ldots\ldots(7.24a)$$

Of course a K^+_{e2} decay mode would still be conceivable for the θ^+ particle but this can proceed here only as a consequence of (7.24*a*), by means of the π–K interaction,

$$\theta^+ \xrightarrow[g_{\pi K}]{} \tau^+ + \pi^0 \rightarrow e^+ + \bar{\nu}. \qquad \ldots\ldots(7.25)$$

Since the matrix element (7.24*a*) is already small and the π–K interaction is only moderately strong, this K^+_{e2} decay (7.25) will be relatively infrequent. For the neutral τ particles, the τ^0 state will have a coupling

$$\tau^0 \rightarrow \overline{\Lambda^0} + N \rightarrow \overline{\Lambda^0} + P + \pi^- \rightarrow e^+ + \bar{\nu} + \pi^-, \qquad \ldots\ldots(7.24b)$$

corresponding to (7.24*a*), the $\bar{\tau}^0$ state being coupled only to the charge-conjugate system $(e^- + \nu + \pi^+)$.

The hyperon decays may be discussed on similar lines (Schwinger 1956). For example $\Lambda_+{}^0$ and $\Lambda_-{}^0$ decays may arise from the sequences

$$\Lambda_+{}^0 \xrightarrow[g_K]{} N + \bar{\theta}^0 \xrightarrow[\lambda]{} N \xrightarrow[g_{\pi\mathcal{N}}]{} P + \pi^-, \qquad \ldots\ldots(7.26a)$$

$$\Lambda_-{}^0 \xrightarrow[g_K]{} P + \bar{\tau}^0 \xrightarrow[g_{\pi K}]{} P + \pi^- + \bar{\theta}^0 \xrightarrow[\lambda]{} P + \pi^-, \qquad \ldots\ldots(7.26b)$$

the final pion–nucleon system necessarily being in a $T = \frac{1}{2}$ state, since the $\Delta T = \frac{1}{2}$ rule is effective in this model. Estimates of the relative $\Lambda_+{}^0$ and $\Lambda_-{}^0$ decay rates, and of the rates for various Σ processes, have been given by Schwinger (1956), and we shall not discuss these in detail here.

A further speculation made by Schwinger is of interest, namely whether a decay interaction

$$\theta_1{}^0 \rightarrow \mu^+ + \mu^- \qquad \ldots\ldots(7.27)$$

might not exist with a matrix element equal to that for the $\theta^+ \rightarrow \mu^+ + \bar{\nu}$ interaction. There is no empirical evidence for this decay mode but less than 1% of $\theta_1{}^0$ particles would be expected to decay in this way. Interaction (7.27) would lead to a τ^+ decay

$$\tau^+ \rightarrow \theta^0 + \pi^+ \rightarrow \mu^+ + \mu^- + \pi^+, \qquad \ldots\ldots(7.28)$$

but again a frequency of only about 10^{-3} relative to $\tau \rightarrow 3\pi$ decay would be expected for this mode. The present absence of evidence for the decay processes (7.27) and (7.28) does not yet exclude the possibility of this muon pair interaction for the $\theta_1{}^0$ particle. The point of interest is that this interaction would allow the transformation of a $\theta_1{}^0$ particle to a virtual μ^+ and μ^- which may then annihilate, the net result being simply the disappearance of a $\theta_1{}^0$ particle, just the mechanism of Schwinger's model for weak decay processes. The expression given for the parameter λ by this mechanism is divergent; the value given by the $\theta_1{}^0$ lifetime is obtained if the divergent integral is cut off for virtual momenta above 4 or 5 nucleon masses.

§ 8. ANOMALOUS EVENTS

The vast majority of events observed to result from high-energy nuclear interactions have allowed a direct interpretation in terms of the well-established interactions of K particles and hyperons, as discussed in the preceding chapters. However, it is important to emphasize the existence of several events for which no satisfactory interpretation has been found possible in terms of the known particles. From certain of these events, it appears probable that at least one 'heavy K particle' (baryon number $B = 0$) or 'super-hyperon' may exist although the properties of this particle are not at all clearly indicated by the present evidence. Almost all of these anomalous events have resulted from stars produced by energetic cosmic-ray particles, for which energies are possible greatly exceeding those available for machine-accelerated particles at present.

8.1. *Unstable Nuclear Fragments with Anomalous Q Values*

Among the very considerable number of unstable fragments which have been observed by various groups, there have been several events in which the decay of the fragment gave a visible energy release exceeding the greatest value possible for decay of a Λ^0 hypernucleus. The clearest of these events has been described by

Fry and Swami (1954). This fragment had charge Z_f exceeding 2 but definitely less than 5, and gave rise to a four-pronged star. One of these prongs was due to a 14 MeV alpha particle, and two other prongs were due to nucleonic particles of charge $Z = 1$ whose energies were 2·7 MeV and approximately 20 MeV if they were assumed to be protons (and correspondingly greater if they were deuterons or tritons). The fourth prong could represent a π^- meson (of 45 ± 18 MeV) or a proton (of 300 ± 125 MeV); in either case the observed energy release is considerably in excess of that available in Λ^0 decay, namely 37 MeV for mesonic decay and 176 MeV for non-mesonic decay. However, charge $Z_f \geqslant 5$ would be necessary for this fragment (contrary to observation) if the fourth prong were a proton, so it seems likely that this prong represents a π^- meson. With a mesonic decay, the visible release of kinetic energy is still at least about $80(\pm 20)$ MeV and there are necessarily neutral secondaries also resulting from the decay.

Fry, Schneps and Swami (1955 c) have also described an energetic non-mesonic decay of a fragment. The secondaries included a grey track representing a proton of 185 ± 20 MeV (or possibly a K particle of about 100 MeV), the total energy release being at least 216 MeV. One further energetic fragment has been reported by Castagnoli, Cortini and Franzinetti (1955). Its three-pronged star consisted of an energetic helium nucleus and two slow particles, the visible energy release being of order 140 MeV; however, the high momentum of the helium nucleus (~ 850 MeV/c) must be balanced by the emission of neutral particles of rather high energy, so that it is difficult to account for the event with an energy release less than about 300 MeV.

These energetic fragments cannot be attributed to Σ hypernuclei, for these can have the necessary stability only in certain exceptional cases (cf. § 6.3). The existence of Ξ hypernuclei stable for times of order 10^{-10} sec would also involve a violation of the principles of the Gell-Mann scheme. However, if it is possible for a K particle to be bound to a nuclear fragment, the above events would be consistent with the decay of K–nuclear fragments, as pointed out by Fry and Swami (1954). The decay processes for θ fragments have been discussed particularly by Pais and Serber (1955). For low θ spin values, they find that, in nuclear matter, the 1-mesonic capture process

$$\theta^0 + \mathrm{P} \rightarrow \pi^+ + \pi^- + \mathrm{P} \rightarrow \pi^+ + \mathrm{N} \qquad \ldots\ldots(8.1)$$

will go at a rate comparable with the normal 2π decay of the bound θ^0 particle; of course, there is in either case a moderately high probability for absorption or interaction of an outgoing pion before escape. At present no energetic fragments exist which are clearly due to decay of a bound K particle. Further, to bind a particle as light as the K particle in a nucleus even as large as carbon or oxygen would require a well depth of about 50 MeV. The evidence on K^+ scattering by nuclei (cf. § 4.5) gives no indication of such a strong attraction; indeed, it now appears that K–nucleus interaction is probably repulsive, corresponding to a positive potential of about 10–20 MeV over the nucleus. The interpretation of these energetic fragments is not at all clear at the present stage.

8.2. *Events with* K *Particle Secondaries*

The Eisenberg particle. Eisenberg (1954) has described an event which appears to represent the decay of a particle to give a K^- secondary. The track of the parent particle (which originated in a moderately large cosmic-ray star) had a deflection of 10° after a path length of 5720 microns and then went on to the end of its range (another 21 900 microns), where a small star was observed. Multiple scattering measurements beyond the point of deflection correspond to a mass $840 \pm 190 m_e$ corresponding to a K particle. A pion was emitted from the five-pronged secondary star, the total visible energy release being 209 MEV, so that this star is quite consistent with being a K^- capture star. However, the scattering measured before the deflection point is about 3·3 times smaller and corresponds to a mass of $3220 \pm 700 m_e$ (Eisenberg 1955). Also there is a significant decrease in the ionization of the track beyond the point of deflection. For these reasons Eisenberg proposed that this deflection represented the decay of a negative hyperon,

$$Y^- \rightarrow K^- + \text{neutral particles.} \qquad \ldots\ldots(8.2)$$

In the rest system of the initial Y^- particle (whose velocity v/c was 0·45), the K^- particle had about 5 MEV kinetic energy. If the neutral particles of (8.2) include a baryon, the Y^- mass necessarily exceeds the Ξ^- mass, and is thus greater than that for any known hyperon.

To ensure the stability of this Y^- particle against fast decay processes, a strangeness assignment of $s = -3$ would appear necessary, and with this an isotopic spin $T = 0$ would be possible, which would correspond to the existence of only a negative variety for this hyperon. The assignment of $s = -2$ or -1 would still allow a rapid γ transition from Y^- to Ξ^- or Σ^- state respectively, without change of strangeness, whilst $s = 0$ would allow rapid processes leading to pions and nucleons. With $s = -3$, it would be natural to assume the decay process

$$Y^- \rightarrow K^- + \Lambda^0 \qquad \ldots\ldots(8.3)$$

for which $\Delta s = +1$; the Y^- mass would then be $3153 m_e$, about 147 MEV above the Ξ^- mass. The production of this particle in nuclear collisions would then require the associated production of three K particles, with a threshold of 6·5 GEV for nucleon–nucleon collisions. The energy of the incident charged particle producing the nuclear interaction in which this Y^- particle originated was about 30 GEV and it has been established by Eisenberg (1955) that at least one of the outgoing prongs probably represents a K particle.

Other suggestions have been made to account for this event, to relate it to known particles or to other anomalous events. Block and King (1955) enquired whether it may not result from a reaction rather than a decay. In this case, with known particles, the incident particle would have to be a Σ^- or Ξ^- particle, the reaction being

$$\Sigma^-(\Xi^-) + N \rightarrow K^- + N(\Lambda^0) + N. \qquad \ldots\ldots(8.4)$$

However, the observations correspond to an incident kinetic energy of only 148 MEV which is insufficient to allow these reactions to proceed. Sachs and Treiman (1955) have suggested that the primary particle may be a bound system of a neutron and a ' heavy K particle ', the K_s^- particle, which they have postulated to account for several events observed by Fry, Schneps and Swami (see below). The difficulties in this proposal will be discussed shortly.

The Wisconsin events. The first of these events consists (Fry, Schneps and Swami 1955 a) of a three-pronged star of which one prong is due to a K meson of

43·8 MeV, the other tracks representing a slow particle of charge $Z = 1$ and an unidentified recoil particle. The K particle was presumed to have negative charge since no secondaries were visible at the end of its range. The primary particle producing this star had a very short track (9 microns) originating in a large cosmic-ray star, but appeared to come to rest. The interpretation of the star is not at all clear. It could represent the nuclear capture of a negative hyperon (or 'heavy K meson') or the decay of a fragment containing a bound hyperon (or 'heavy K meson').

The second event (Fry, Schneps and Swami 1955 b) had some similarity to this event. It consisted of a particle of roughly protonic mass which originated from a cosmic-ray star and came to rest after a path length of 44 microns, giving a K particle secondary of 42 MeV. This K particle gave no visible secondaries at the end of its range and was therefore presumed negative.

Sachs and Treiman (1955) have proposed to interpret these events in terms of a 'heavy K particle', named the K_s^- particle, with strangeness assignment $s = -2$. With the relation (2.23), an isotopic spin assignment $T = 0$ would be appropriate for this particle, as this avoids the assumption of doubly charged K_s particles. Reactions of the type

$$K_s^- + N \begin{cases} \rightarrow \Lambda^0 + K^-, & \quad\ldots\ldots(8.5a) \\ \rightarrow \Xi^- + \pi^0, & \quad\ldots\ldots(8.5b) \end{cases}$$

can then take place with conservation of strangeness and isotopic spin. The two Wisconsin events can then be attributed to the occurrence of this capture reaction (8.5*a*) after the K_s^- particle has come to rest; the K^- energies observed would require a K_s mass of about $1500\, m_e$.

However, Sachs and Treiman then proposed to go further and to account for the Eisenberg particle as due to a similar reaction,

$$K_s^- + N \rightarrow \Sigma^0 + K^- \qquad \ldots\ldots(8.6)$$

occurring for a bound $(K_s^- + N)$ system. Only if these reactions (8.5), (8.6) which conserve strangeness were slow (contrary to the principles of the Gell-Mann scheme) could this system be stable for a time as long as about 10^{-10} sec. In order to achieve this, Sachs and Treiman proposed $T = \frac{1}{2}$ for the K_s particle, a proposal which violates the relation (2.23) established for all other strongly interacting particles. This proposal is equivalent to the introduction of a new quantum number b with the value $b = +1$ for the K_s particle and $b = 0$ for all other strongly interacting particles, and with the property that $\Delta b = 0$ for all fast reactions. This far-reaching assumption rests only on the desire for related explanations of the Eisenberg and Wisconsin events. However, these two types of event can be given (as above) separate explanations within the Gell-Mann scheme, by using values of the available quantum numbers not yet assigned to other particles, so that there is at present no compelling necessity to go outside this well-established scheme. Further evidence on anomalous events of these types will clearly be of the greatest interest.

§ 9. CONCLUSION

9.1. *The Gell-Mann–Nishijima Scheme and Particular Theoretical Models*

In the chapters above it has been clear that our present knowledge of the quantitative properties of K particles and hyperons is in general quite preliminary and may undergo considerable revision as experimental techniques improve and more extensive and detailed data are collected. However, the qualitative features of the strong interactions involving strange particles now appear reasonably clear. In all the reactions which have been established, the sum of the strangeness numbers s for all the particles involved has been conserved through the reaction, the strongly interacting particles being assigned the values of s arising from the Gell-Mann–Nishijima scheme (cf. figure 1). Also certain reactions such as (2.9) which satisfy the Associated Production principle (cf. § 2.1) but which do not allow conservation of strangeness, have not been detected, despite quite sensitive observations (cf. § 3.2). Further, conservation of strangeness is definitely violated in all of the slow decay processes which result in strongly interacting particles, a change $\Delta s = \pm 1$ being the rule for these decays. The principle that interactions can be strong only if they allow conservation of strangeness now appears rather well established.

The second aspect of the Gell-Mann–Nishijima scheme is the hypothesis of charge independence. This is suggested by the observation that strange particles, as well as the nucleons and pions, appear in multiplets, the isotopic spin T assigned to a particular multiplet being given by its multiplicity $2T+1$. The direct experimental evidence that charge independence does hold for the strong interactions involving strange particles is still very limited. Evidence on light hypernuclear states (cf. § 6.2) is consistent with the assignment $T=0$ for the Λ^0 particle and certainly suggests that the Λ^0 particle is charge symmetrical to itself; the data on the K$^-$–proton capture reaction is at present consistent with the inequalities following from charge independence. Certain experiments which will allow a direct test of the hypothesis of charge independence have been mentioned in §§ 2.5 and 4.7, and these experiments will undoubtedly be carried through over the next year or two. From our knowledge of charge independence for the pion–nucleon interaction and the appearance of similar multiplets for the strange particles, it is our natural expectation that charge independence will be found valid, at least to a good approximation, and this has generally been assumed.

Our present semi-quantitative description of strange-particle phenomena therefore contains two elements: (*a*) the existence of multiplets to each of which an isotopic spin T is assigned, the component T_3 running from $-T$ for the particle of least positive charge to $+T$ for the particle of most positive charge in the multiplet, and (*b*) the displacement of the centre of each multiplet, relative to the nucleon multiplet for hyperons or relative to the pion multiplet for K particles. Counting this displacement as positive towards more positive charge states, it is directly related to the strangeness, being exactly $s/2$. For the known multiplets, these parameters s and T are not completely independent, in view of the relation

$$Q = T_3 + \tfrac{1}{2}(s+B) \qquad \ldots\ldots(9.1)$$

which connects the charge of each particle with its isotopic spin component T_3, strangeness s and baryon number B. The combination

$$U = s + B \qquad \ldots\ldots(9.2)$$

appears to have useful properties since, for the known particles, its values are limited to 0 (pions, Λ and Σ particles), $+1$ (K particles and nucleons) and -1 ($\bar{\text{K}}$ particles and Ξ particles). Also integral isotopic spin holds for the particles with even

$U(=0)$, half-integral isotopic spin for particles with odd $U(=\pm 1)$; with the relation (9.1) this is related to the quantization of charge in units of the electron charge.

With these facts before us, there has naturally been a good deal of speculation with the aim of finding a deeper theoretical basis for the interpretation of the existence and properties of strange particles, to understand (for example) the origin of the relation (9.1) and why just these multiplets should appear with the displacements $s/2$ observed. An early attempt along these lines was made by Pais (1954), who proposed the introduction of two isotopic spin vectors, $\mathbf{K}$ as well as $\mathbf{T}$. These two isotopic spin vectors are just sufficient for the specification of a four-dimensional isotopic-spin space. Conservation of both K^2 and T^2 as well as their components K_3 and T_3 would correspond to the requirement that all strong interactions be invariant with respect to rotations in this four-dimensional isotopic-spin space, and this was the hypothesis proposed by Pais. This assumption implies the existence of double-degenerate multiplets, with multiplicity $(2T+1)(2K+1)$. The strangeness s of a particle is then expressed in terms of the third component K_3 of this second isotopic-spin vector. For the nucleon, $T=\frac{1}{2}$ so that $K=0$ gives the observed multiplicity; this requires that $s=2K_3$ and this leads to the general relation

$$Q = T_3+K_3+\tfrac{1}{2}B. \qquad \ldots\ldots(9.3)$$

Then the multiplets (T, K) must be chosen such that $T+K$ is half integral for hyperons and integral for K particles.

The difficulties of the proposal may be illustrated by reference to the Λ^0 particle. Since $T=0$ for this particle, $K=\frac{1}{2}$ is the least value of K possible. However, this requires a doublet Λ state, a Λ^0 state with $K_3=-\frac{1}{2}$ and a Λ^+ state corresponding to $K_3=+\frac{1}{2}$ (which means strangeness $s=+1$). It is quite certain (from the absence of absorptive K$^+$–nucleon reactions, for example) that there does not exist a Λ^+ particle of the same mass as the Λ^0 particle and with $s=+1$. This discrepancy can be avoided if this Λ doublet is split with respect to K_3 by strong interactions, so that the unobserved Λ^+ state has a mass exceeding m_n+m_K; in this case the Λ^+ particle would have a rapid $\Lambda^+\rightarrow P+K^0$ decay through strong interactions and would be unobservable due to its very short lifetime. However, this assumption means that K^2 will not be even approximately conserved and that the strong interactions do not have the four-dimensional isotopic-spin invariance which was proposed, and this model, although still possible, no longer appears attractive.

Salam and Polkinghorne (1955) have suggested that, in this four-dimensional isotopic-spin scheme, K_3 may be identified with $\frac{1}{2}U=\frac{1}{2}(S+B)$ rather than $\frac{1}{2}S$, so that

$$Q = T_3+K_3. \qquad \ldots\ldots(9.4)$$

In this case the Λ^0 particle corresponds to $T=K=0$, a singlet state, so that this proposal is free from the difficulty mentioned above. However, with $T=\frac{1}{2}$ for the neutron–proton doublet, $K_3=+\frac{1}{2}$ is appropriate for the nucleons. In order to account for the remaining particles of this $T=K=\frac{1}{2}$ nuclear multiplet, Salam and Polkinghorne propose that the $K_3=-\frac{1}{2}$ component of this multiplet be identified with the Ξ particles. Again, the large mass difference observed between these two components of the $T=K=\frac{1}{2}$ multiplet implies that the strong interactions must have very considerable dependence on K_3 and therefore do not possess the four-dimensional isotopic-spin invariance appropriate to the framework of this model.

A simpler scheme has recently been discussed by d'Espagnat and Prentki (1956 a) and by Racah (1956). These authors consider the usual three-dimensional isotopic-spin space but require that the strong interactions remain invariant under reflections as well as rotations in this space. This new symmetry property leads to the introduction of an isotopic parity (which we denote by v) describing how the wave function for each particle behaves for a reflection of the isotopic-spin space. For pions and Λ, Σ hyperons, isotopic parity $v = +1$ is assigned. The nucleons, Ξ particles and K, $\bar{\mathrm{K}}$ particles each have $T = \frac{1}{2}$ and must therefore be described by spinor wave functions in isotopic-spin space, just as the nucleon spin state is described by a spinor wave function in ordinary spin-space. The nucleons and K particles are assumed to have isotopic parity $+i$ (which corresponds to spinors of the first kind), whilst the Ξ particles and K particles are distinguished from these particles by the assignment $v = -i$ (which corresponds to spinors of the second kind). These assignments of isotopic spin parity may be summarized by the expression

$$v = (i)^U = (i)^{s+B}. \qquad \ldots\ldots(9.5)$$

The invariance of strong interactions under reflections in isotopic-spin space implies the conservation of isotopic parity through all fast processes. Equating the product of the isotopic parities for the initial particles with that for the final particles then gives the relation

$$(i)^{\Sigma_{\mathrm{i}}(s+B)} = (i)^{\Sigma_{\mathrm{f}}(s+B)}, \qquad \ldots\ldots(9.6)$$

where the sums over i and f refer to the initial and final particles. From this it follows that

$$\sum_{\mathrm{f}}(s+B) - \sum_{\mathrm{i}}(s+B) = 4n, \qquad \ldots\ldots(9.7)$$

where n is an integer (since $(i)^{4n} = 1$). D'Espagnat and Prentki next assume: (i) that the basic strong interactions are all of the Yukawa type, $a \longleftrightarrow b+c$, involving only three particles, (ii) that only particles with $U = -1, 0, +1$ exist, corresponding to the situation for the well-established particles.

Since isotopic parity is naturally conserved for each basic interaction, an equation of the form (9.7) will hold for each of these. With conditions (i) and (ii), the left-hand side of (9.7) can have a magnitude only less than or equal to 3, so that $n = 0$ must hold for each basic interaction. However, any fast reaction results from a sequence of (virtual) basic interactions, so that generally, for any process,

$$\sum_{\mathrm{i}}(s+B) = \sum_{\mathrm{f}}(s+B), \qquad \ldots\ldots(9.8)$$

i.e. the sum of s and B is conserved. Since rigorous conservation of the baryon number B is naturally imposed, it follows that strangeness conservation must hold for all the strong interactions of this model, provided conditions (i) and (ii) and the above assignments of isotopic parity v are assumed.

In this model, charge is *defined* by the relation

$$Q = T_3 + \tfrac{1}{2}(s+B). \qquad \ldots\ldots(9.9)$$

From this viewpoint charge conservation necessarily holds for strong interactions in consequence of their invariance under rotations and reflections in isotopic-spin space. However, conservation of isotopic parity and of T_3 are known to hold only for the strong and electromagnetic interactions, being clearly violated in the weak decay processes, whereas charge conservation is well known to hold rigorously. The assumptions of this model do not lead automatically to a rigorous conservation of charge, but this must be imposed as a separate *ad hoc* requirement after definition (9.9). It should be added here that this criticism applies equally to the two models described above.

With the condition (ii) above, all sets of isotopic spin quantum numbers available for strange particles in the d'Espagnat–Prentki model have been assigned to known particles, assuming that doubly charged particles are forbidden. If further strange particles exist, this model would be shown inadequate and new quantum numbers would have to be introduced for their description. Further evidence on the anomalous events discussed in § 8, which suggest the existence of heavier unstable particles, will be of particular importance for this model.

The simplest set of K particle interactions of the Yukawa type which satisfy the requirements of d'Espagnat and Prentki are (assuming the θ particle to have an s-wave interaction with Λ and Σ particles) as follows :

$$g_1\{(\theta\bar{\mathcal{N}})\,\Lambda+\text{c.c.}\}+g_2\{(\theta\boldsymbol{\tau}\bar{\mathcal{N}})\,.\,\boldsymbol{\Sigma}+\text{c.c.}\} \\ +g_3\{(\bar{\theta}\eta\bar{\Xi})\,\Lambda+\text{c.c.}\}+g_4\{(\bar{\theta}\boldsymbol{\tau}\eta\bar{\Xi})\,.\,\boldsymbol{\Sigma}+\text{c.c.}\}. \qquad \ldots\ldots(9.10)$$

In (9.10), each field function has been denoted by the symbol characteristic of the corresponding particle, for example the first term describes a basic interaction $\mathcal{N}\to\Lambda+\theta$ with a matrix element g_1; c.c. denotes the complex conjugate term in each case and the scalar factor η is defined by the requirement that $\eta\Xi$ be a spinor of the first kind (Salam 1956). If invariance under parity conjugation is appropriate, then $\Lambda, \boldsymbol{\Sigma}$ in (9.10) are to be replaced by $\Lambda_+, \boldsymbol{\Sigma}_+$ and the interaction will contain a further four terms obtained from (9.10) by replacing θ by τ and $\Lambda_+, \boldsymbol{\Sigma}_+$ by $\Lambda_-, \boldsymbol{\Sigma}_-$; naturally these additional terms introduce no further coupling parameters.

Schwinger (1956) has recently discussed a theory whose K particle nuclear interaction is obtained from (9.10) by the specialization

$$g_1 = g_2 = g_3 = g_4 = g_K, \qquad \ldots\ldots(9.11)$$

leading to a one-parameter interaction. Schwinger points out that this interaction then has a certain symmetry between Λ and $\boldsymbol{\Sigma}$ fields, namely that it remains invariant when Λ and $\boldsymbol{\Sigma}$ fields are mixed by a linear transformation (formally identical with a Lorentz transformation in the $(\Lambda, \boldsymbol{\Sigma})$ space), provided certain corresponding transformations are simultaneously applied to each of the $\mathcal{N}$, the Ξ and the θ fields. This symmetry means that if Λ and Σ particles had the same bare mass in the absence of all interactions, then Λ and Σ masses would continue to be equal when this K interaction is effective. Similarly, the interaction has a certain symmetry for interchange of $\mathcal{N}$ and Ξ fields, with the result that a mass equality between $\mathcal{N}$ and Ξ would be maintained by the K interaction.

Schwinger introduces only two other strong interactions, the π–$\mathcal{N}$ interaction with coupling parameter $g_{\pi\mathcal{N}}{}^2 \sim 15$ and the π–K interaction between τ and θ discussed in § 7.4. The mass spectrum observed for the hyperons is now attributed to these pion interactions which will split the $(\Xi, \mathcal{N})$ mass equality since only the nucleons have a direct pion interaction and which will also give a smaller higher-order splitting between Λ and Σ masses. The splitting of this four-dimensional symmetry between $(\Xi, \mathcal{N})$ and (Λ, Σ) fields by the π–$\mathcal{N}$ and π–K interactions is regarded as analogous to the splitting of the three-dimensional symmetry of the usual isotopic-spin space by the electromagnetic interactions, which leads, for example, to a mass separation between $\pi^{\pm}$ and π^0 states. This analogy is much weakened by the fact that the π–$\mathcal{N}$ interaction is very strong, whilst the electromagnetic interaction is quite weak. The pion interactions will strongly distort the four-dimensional symmetry of the K interaction (9.11), in contrast with the rather weak electromagnetic distortion of the three-dimensional isotopic-spin symmetry in the pion–nucleon interaction. In fact, until methods are developed for the reliable calculation of the consequences of strong interactions, the existence of a symmetry such as (9.11) between these interactions could not be recognized even if very detailed empirical data relevant to all these interactions were available, since their distortion due to the π–$\mathcal{N}$ interaction would still be unknown.

The formal development of Schwinger's model is quite different in detail from that of d'Espagnat and Prentki. For example, although invariance under reflections in isotopic-spin space is required for strong interactions, the operation corresponding to this reflection is here expressible in terms of the product of charge conjugation and of a rotation in isotopic-spin space, and therefore does not lead to any new quantum number. However, the physical content of the theory is contained in the symmetry (9.10) required for the K interaction and in the assumption that the π–$\mathcal{N}$ and π–K interactions are the only other strong interactions.

In our present state of empirical knowledge and with our inability to calculate reliably for the strong coupling theories appropriate to these strong interactions, it is clear that these theoretical models are to be regarded as only exploratory at this stage. In general the only features of the models which can be tested experimentally are those depending on the conservation laws which are implied by the interactions considered, but which are necessarily just the conservation laws of the Gell-Mann–Nishijima scheme. However, these models have definite value in helping us to realize some of the possibilities which may underlie the problems before us.

9.2. *Outlook*

There are many aspects of the character of strange particles and their strong interactions which are yet to be settled by experiment before any theoretical model can be regarded as firmly based. The lack of direct evidence concerning the charge independence of these interactions has already been mentioned above. There is also rather little evidence bearing on the spins and parities of these particles. Their direct measurement appears a more difficult proposition than was the case for the pion, for a pion can undergo absorptive nuclear reactions which lead to final states of known properties and often with special symmetries whereas, with the ' Associated Production ' character of strong interactions, nuclear absorption of a strange particle leads to the formation of another strange particle whose properties are uncertain. In due course, experiments in which cross sections are measured at a given centre-of-mass energy for the reactions $A+B \longleftrightarrow C+D$ proceeding from the left and from the right (e.g. for the reactions $K^- + D \longleftrightarrow \Sigma^- + P$) will become possible. From this data, arguments of detailed balancing will lead to a knowledge of the ratio $(2S_\Lambda+1)/(2S_C+1)$, B and D being assumed known particles. Only quite low spin values are reasonable for the strange particles, from arguments given in the preceding chapters, so that only certain discrete values are possible for this ratio; as pointed out by Lee (1955), sufficiently accurate measurements may allow the spins of both strange particles to be determined (or at least limited to certain joint values). However, the situation is naturally much simpler if the K particle is assumed to have spin zero (either parity). In this case, as Adair (cf. § 3.1) has emphasized, some definite statements can be made about the angular correlations in the decay of Λ and Σ particles produced in the forward or backward direction from the $(\pi+\mathcal{N})$ reactions. Further statements may be made on these angular correlations, irrespective of the Λ or Σ direction, provided the pion energy is sufficiently close to threshold for the final $(\Lambda+K)$ or $(\Sigma+K)$ system to have s-wave orbital motion. These statements are especially valuable since the observation of isotropic decay would then correspond uniquely to spin $\frac{1}{2}$ for the hyperon. Experimental results on these angular correlations should soon be available with good statistics from bubble chamber studies now in progress.

The relationship between τ and θ particles is especially obscure at present. If parity conservation fails in weak decay processes (cf. § 7.3), which would provide the most direct interpretation of the evidence summarized in § 7, this may be difficult to establish. There are characteristic angular correlations, impossible with parity-conserving decay, which can appear in hyperon decays, and a careful search for such effects in a wide range of processes will be important; however, their absence would not provide an argument against parity non-conservation. Non-conservation of parity in weak decay processes would mean that no information on the parities of strange particles could be obtained from the angular correlations in their decay (e.g. the conclusions on the parity of ${}_\Lambda H^4$ in § 6.2 would no longer hold). In fact, with strong interactions which always involve two strange particles, parity would have no physical meaning for an isolated Λ, Σ or K particle; only relative parities between a pair of these particles would then be physically significant since the forms permissible for the strong interactions depend only on the total parity of pairs of these particles. On the other hand, the existence of parity-conjugation invariance may well be difficult to demonstrate. As pointed out in § 7.3, correlated observations on the decay of both strange particles produced in association will bear very directly on this question. For example, the existence of a Λ^0 parity doublet requires that the decay curve of Λ^0 particles for which accompanying $\theta_1{}^0$ decay is observed should differ from the decay curve for those Λ^0 particles for which $\theta_1{}^0$ decay is not observed (unless $\Lambda_1{}^0$ and $\Lambda_2{}^0$ particles also have the same lifetime, an unlikely coincidence); similarly, the branching ratios for K^+ decay accompanied by $\theta_1{}^0$ decay in pair production processes such as

$$P+P \rightarrow K^+ + K_1{}^0 + P + N$$

should generally be different from those observed for K^+ particles irrespective of their accompanying particle. Experiments of this kind are clearly difficult and will make slow progress. At the present time the major difficulty with the parity conjugation hypothesis is the τ–θ lifetime identity, which is viewed only as an approximate equality; although these lifetimes are now quite well known, still more accurate measurements are desirable to establish this lifetime equality as closely as possible.

The behaviour of excitation functions for strange particle production near threshold, the energy and angular distributions of scattering and reaction cross sections for $K^\pm$, Λ and Σ particles, all of these will be of importance in establishing the qualitative characteristics of the form of the K particle interaction with nucleons, for example the angular momentum and isotopic spin states in which the K particle nuclear interaction is strongest. At best there are only quite preliminary data available on these processes at present. However, the development of bubble chambers of various materials and the achievement of almost pure K^+ and K^- beams from the existing machines will lead to a rapid improvement in the accuracy and reliability of our knowledge of the fundamental processes. Also the electromagnetic interactions of strange particles may now be investigated, for electrons and photons of up to 1·2 GeV energy are now available in several centres. Further proton accelerators at present under construction will allow experiments with proton beams of much higher energy (up to ~30 GeV) or of much higher intensities (by a factor ~100 at ~3 GeV) than existing machines to be made within the next five years or so.

The extension of strange-particle research to still higher energies is especially important, in order to establish whether or not there exist still further, heavier strange particles, as certain anomalous events suggest at present. Our knowledge of the existence of the Ξ^- particle has been of great importance, yet only seven examples of decay are known, all from cosmic-ray experiments. As shown by the anomalous events discussed in § 8.2, all of which resulted from cosmic-ray exposures, cosmic-ray experiments (especially at high altitudes) will continue to be of importance for many years for exploration of the possibilities for phenomena involving new particles or higher energies. It is entirely possible that strange particles with very short lifetimes may also exist; this will certainly be the case if there are heavier particle states with the same charge, baryon number and strangeness as known lighter particles. One example is already known, namely the Σ^0 particle whose lifetime is believed to be in the region of 10^{-20} sec; the width corresponding to this lifetime is of the order of 0·1 MeV so that the Σ^0 particle represents a well-defined state. For this example the lifetime is unusually long, from the fact that the mass difference between Σ^0 and Λ^0 particles is smaller than the pion mass, which means that the only decay process allowed by the conservation laws is an electromagnetic decay. If the Σ^0 particle were more than 30 MeV heavier than it is, then a rapid pion decay to the Λ^0 particle would be possible with a lifetime at least 100 times shorter than the observed Σ^0 decay; the Σ^0 state would then have appeared as a resonance in $\Lambda^0+\pi^0$ scattering, whose presence could be detected only from the study of the Q value distribution between Λ^0 and π^0 particles resulting from strange-particle production processes (obviously a difficult proposition). If the Σ^0 particle were still about another 200 MeV heavier than this, fast K particle decay $\Sigma^0 \rightarrow K^- + P$ would be possible and the Σ^0 particle would have appeared also as a resonance in $K^- + P$ scattering. A resonant scattering state of this kind is already known in pion physics, for the pion–nucleon system has a strong resonance (the $T = \frac{3}{2}, j = \frac{3}{2}$ isobar state, with half width about 70 MeV) at about 140 MeV pion kinetic energy (c.m. system); in addition to its direct observation in pion–nucleon scattering and pion photoproduction, this resonance also becomes apparent in the $Q(\pi^+, P)$ distribution for pion production in proton–proton collisions. The investigation of similar resonant states for strange-particle systems will be of comparable importance for the detailed structure of the theory of strange particles and their interactions.

Only one thing certain can be said. Many fascinating years are immediately ahead of us, during which these very subtle and interesting phenomena will be explored and interpreted, bringing us continually nearer to some understanding of the deepest levels of the structure of matter which have yet become apparent to us.

ACKNOWLEDGMENTS

The above article has been written mainly during a period when the author was on leave from the University of Chicago, still at the Department of Mathematical Physics of the University of Birmingham. He has appreciated the hospitality of the University of Birmingham, the facilities made available to him at the Department of Mathematical Physics, and the warm interest shown by Professor R. E. Peierls and other members of this Department in the lecture course from which this article originated. Thanks are due especially to Dr. W. O. Lock, Dr. J. B. Woodward and Mr. H. Burkhardt for their critical reading of the manuscript and to Professor R. E. Peierls for his comments on certain sections. The author has found especially valuable the careful summary by Professor R. W. Thompson (Thompson 1956) of the experimental situation up to 1955 and the thoughtful review of the emulsion evidence on the nuclear capture of K$^-$ particles by S. Goldhaber (1956). The Rochester Conference reports have provided a source of information and stimulation beyond compare.

Finally, the author would like to express his gratitude for the permission to reproduce many of the diagrams of this article from various published papers, which has been granted him by the authors and publishers of these papers. Thanks are also due to Dr. V. Telegdi for discussions concerning the Appendix added in proof.

APPENDIX

(*added May* 1957)

INVARIANCE PRINCIPLES VIOLATED IN THE BETA-DECAY AND THE π–μ–e PROCESSES

During the summer of 1956, as discussed in §7.2, Lee and Yang (1956 b) proposed a number of experiments which could establish whether or not parity was conserved in the weak decay processes which are most familiar to us, nuclear beta-decay and the π–μ and μ–e decay processes. They pointed out that previous experiments on these decay processes gave no information on this question. A number of the experiments proposed have now been carried through and have shown us conclusively that at least two basic invariance principles well established for strong interactions, invariance under space reflection and invariance under particle–antiparticle interchange (charge conjugation, cf. §6.2), are not satisfied by these weak decay interactions.

For nuclear beta decay, this was first established by Wu *et al.* (1957) from observations on the beta-decay of partially oriented ^{60}Co nuclei, for which a polarization of 60% had been achieved by the Rose–Gorter method, using the low-temperature facilities of the National Bureau of Standards. In this experiment Wu *et al.* found that, in the velocity range considered ($v_e/c \simeq 0{\cdot}65$), the emission of beta particles from the polarized nuclei into the backward hemisphere (relative to the polarization direction) was about 50% more intense than the emission into the forward hemisphere. Such a forward–backward asymmetry can, of course, occur in the decay of a nuclear state of definite parity only if the decay interaction includes terms which reverse the parity of the system as well as terms which conserve total parity, since it arises only from the interference between final states of opposite parity. In fact, for an allowed Gamow–Teller transition with spin change $J \to J-1$ (such as the beta transition of ^{60}Co, where $J=5$), which can be produced only by the tensor (**T**) and axial vector (**A**) beta interactions, the coefficient A in the angular distribution

$$\left(1 + A\frac{v_e}{c}P\cos\theta\right) \qquad \ldots\ldots(\mathrm{A1})$$

given by Lee and Yang (1956 b) for electrons (velocity v_e) emitted in an allowed transition from polarized nuclei of polarization $P = \langle J_3 \rangle / J$ reduces to

$$A = 2\frac{[\mathscr{R}(c_T{}^+ c_T{}^{-*} - c_A{}^+ c_A{}^{-*})] + (Z\alpha/p_e)[\mathscr{I}(c_A{}^- c_T{}^{+*} + c_A{}^+ c_T{}^{-*})]}{|c_A{}^+|^2 + |c_A{}^-|^2 + |c_T{}^+|^2 + |c_T{}^-|^2}, \qquad \ldots\ldots(\mathrm{A2})$$

$c_T{}^+, c_A{}^+$ being the interaction constants for the tensor and axial vector interactions which conserve parity, $c_T{}^-, c_A{}^-$ being those for the interactions which reverse parity. In (A2), Z denotes the nuclear charge, α is 1/137 and p_e denotes the momentum of the beta particle.

The experimental results quoted above for ^{60}Co correspond to a coefficient $A \simeq -1{\cdot}0$. This value has been roughly confirmed by several experiments which measure other polarization phenomena associated with ^{60}Co beta decay. For example, with parity non-conservation in beta-decay, the beta particles emitted from *unpolarized* nuclei will generally have a longitudinal polarization, i.e. a spin axis preferentially parallel or antiparallel to their direction of emission. In ^{60}Co decay, their degree of longitudinal polarization is determined by the same parameter A given in (A2), the electron polarization parallel to its motion being Av_e/c. Of course the observation of a longitudinal electron polarization is somewhat difficult, the well-known Mott scattering method being effective only for the transverse component of the polarization. In these preliminary experiments various methods have been used. Frauenfelder *et al.* (1957) have measured the polarization by turning the path of the beta particles from ^{60}Co through 90° by an electrostatic deflection—this results in a beam of transversely polarized electrons whose polarization may then be measured by the Mott scattering method, the measurement of left–right asymmetry in their subsequent scattering by a thin foil. Goldhaber, Grodzins and Sunyar (1957) have measured the longitudinal polarization by observing the circular polarization of the energetic bremsstrahlung produced by high-energy beta particles emitted from ^{90}Y. Page and Heinberg (1957) have measured longitudinal polarization of positive beta particles by making use of some spin dependent effects in positron annihilation. Again Schopper (1957) has determined the circular polarization of gamma rays emitted following an allowed beta transition, which also depends on the same parameter A for a pure Gamow–Teller transition. For a mixed Fermi and Gamow–Teller transition, experiments of the type just mentioned here give additional information concerning the coupling parameters, since the various experiments measure somewhat different combinations of the Fermi and Gamow–Teller coupling parameters. In the case of pure Fermi transitions, electron polarization experiments measure combinations of the Fermi couplings which are quite inaccessible to the nuclear polarization experiments, since the Fermi interactions do not couple the electron–neutrino motion with the nuclear spin. However, at the present stage the art is still in a preliminary state, and there is still some conflict between several of the experimental results already available.

In the expression (A2) for the coefficient A, the first bracket may be shown to vanish if the beta interaction remains invariant under the operation of charge conjugation (which we shall denote here by C). This follows as a special case of a general theorem by Lee, Oehme and Yang (1957) to the effect that no manifestation of parity non-conservation can ever be detected if the weak interaction has invariance under C and if the forces between the final particles in the decay can be neglected. On the other hand, if the beta interaction is invariant under time reversal (T), then the interaction constants $c_T^{\pm}, c_A^{\pm}$ are all real (Pauli 1955) and the second bracket of (A2) will vanish.

Now it is already known from electron–neutrino angular correlation experiments in ^{6}He decay that the axial vector interactions are relatively the weaker,

$$(|c_A^+|^2+|c_A^-|^2)/(|c_T^+|^2+|c_T^-|^2)\lesssim 1/3.$$

From this it can readily be seen that such a large value of A as that observed for ^{60}Co cannot be obtained from the second bracket of (A2) alone. Indeed, the measurements of A by Wu *et al.* (1957) and by Frauenfelder *et al.* (1957) show that A is essentially independent of electron velocity v_e so that the second bracket of (A2) is necessarily small—of course, this may simply result from the smallness of axial vector relative to tensor interactions and does not necessarily imply their invariance with respect to time reversal T. However, the presence of the first bracket of (A2) is necessary to account for the observed value of A and this implies that the beta interaction is not invariant under C. The observed value of A also requires that c_T^+ and c_T^- be of comparable magnitude and of opposite sign,

$$c_T^- \simeq -c_T^+. \qquad \ldots\ldots(\text{A3})$$

Although the beta interaction is not invariant under space reflection P or under charge conjugation C, it is still possible that it may be invariant under the product CP of these operations carried out jointly. Now Lee, Oehme and Yang (1957) have restated the Lüders theorem (cf. § 6.2) in the following form, that the Hamiltonian of any Lorentz-invariant, local theory in which the usual connection between spin and statistics holds is necessarily invariant with respect to the product of the three operations C, P and T, taken in any order. From this it follows that invariance of the interactions under the product operation CP implies that the interactions are invariant under time reversal T, and conversely, that invariance under time reversal T requires invariance of the interactions with respect to CP. This CP invariance would correspond to an attractive situation in that an overall symmetry in space–time would still hold valid. With this invariance, physical phenomena would still satisfy the same laws after a mirror reflection of space in which right-handed and left-handed systems are interchanged, provided that at the same time all particles are replaced by their anti-particles. For example, a ^{60}Co nucleus emits electrons which spin with a left-handed sense. With CP invariance, an anti-^{60}Co nucleus composed of antiprotons and antineutrons will decay with emission of positrons with a right-handed sense of spin. It is then clear that the properties of this beta-decay process could be used to provide a physical definition of a left-handed as opposed to a right-handed system of axes only if there were some absolute means to distinguish positive charge, to establish that the decaying nucleus is ^{60}Co rather than anti-^{60}Co, without reference to special properties of the universe in our locality. In this way, we see that the definite sense of spin for beta particles emitted from radioactive nuclei is related to the fact that the world in our locality contains a great predominance of protons of one charge sign over protons of the opposite charge, and is not necessarily due to any asymmetry in space–time itself.

Another attractive situation consistent with this result (A3) has been pointed out by Landau (1957) and by Lee and Yang (1957). With the equality (A3), the tensor interaction takes the form

$$(\bar{\psi}_P \sigma_{\alpha\beta} \psi_N)(c_T^+ \bar{\psi}_e \sigma^{\alpha\beta} \psi_\nu + c_T^- \bar{\psi}_e \sigma^{\alpha\beta} \gamma_5 \psi_\nu)$$
$$= c_T^+ (\bar{\psi}_P \sigma_{\alpha\beta} \psi_N)(\bar{\psi}_e \sigma^{\alpha\beta}[1-\gamma_5]\psi_\nu). \qquad \ldots\ldots(\mathrm{A4})$$

Now for particles of zero rest mass, the operator γ_5 is equivalent to

$$\gamma_5 = \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E_\nu} = \pm\sigma_3,$$

the upper or lower sign being taken for neutrino or antineutrino respectively, where σ_3 is the spin operator parallel to the momentum of the particle. Thus the wave function $(1-\gamma_5)\psi_\nu$ is zero except for neutrinos which spin with left-handed sense, or for antineutrinos spinning with a right-handed sense. This provides a covariant neutrino theory, in which there are only two states for given momentum, a left-handed neutrino state and a right-handed antineutrino state. This two-component neutrino theory exists only if the neutrino mass is precisely zero. The existence of this theory was first noted by Weyl (1929), and it was Salam (1957) who first emphasized that the condition that all interactions be invariant under the replacement $\psi_\nu \to -\gamma_5\psi_\nu$, which ensures that $m_\nu = 0$ to all orders, would require parity non-conserving beta-decay interactions of the type displayed in (A4). It is clearly possible for the interactions of this theory to be CP invariant since the operation P replaces left-handed neutrinos by right-handed neutrinos, whilst C replaces right-handed neutrinos by right-handed antineutrinos, so that CP replaces left-handed neutrinos by right-handed antineutrinos and vice versa.

The empirical situation concerning invariance under time reversal T (and therefore under CP) is not at all clear at present. There is certainly no clear evidence to indicate that this invariance property does not hold for the beta interaction, although most of the experiments completed have not been sensitive to violation of this property. One fundamental experiment which would bear very directly on this question has been proposed by Jackson, Treiman and Wyld

(1957 a), who suggest measuring the coefficient of $\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{p}_e\times\mathbf{p}_\nu$ in the decay of polarized neutrons, i.e. the up–down asymmetry of the normal to the neutron decay plane relative to the neutron polarization. Such an asymmetry can exist only if the beta-interaction is not invariant under T; this is clear from the fact that the product $\langle\boldsymbol{\sigma}\rangle\cdot\mathbf{p}_e\times\mathbf{p}_\nu$ reverses under time reversal, since the neutron spin $\langle\boldsymbol{\sigma}\rangle$ and the electron and neutrino momenta each reverse under time reversal. The explicit dependence of the coefficient of this term on the beta-decay coupling parameters is proportional to the following expression

$$M_F M_{GT}\mathscr{I}(c_S^+ c_T^{+*} + c_S^- c_T^{-*} - c_V^+ c_A^{+*} - c_V^- c_A^{-*}). \qquad \ldots\ldots(\mathrm{A5})$$

This important experiment to measure the coefficient (A5) is at present in progress at Chalk River and at the Argonne National Laboratory.

The other weak interactions for which parity non-conservation has been clearly established are those giving rise to the π–μ and μ–e decay processes. For brevity, this situation will be discussed only in terms of the two-component neutrino theory—a more complete discussion has been given by Bouchiat and Michel (1957). From angular momentum conservation in the decay of the pion (spin zero), it is clear that the muon must be emitted spinning in the same sense as the neutrino (or antineutrino) emitted in the decay, i.e. the muon must be emitted with complete longitudinal polarization. When this muon comes to rest and undergoes beta-decay, the electrons emitted will generally have forward–backward asymmetry relative to the direction of polarization of the muon, provided both P and C are violated in the muon–beta interaction (the forces between the final particles being negligible). The existence of such an asymmetry has been established by Garwin, Lederman and Weinrich (1957), by Friedman and Telegdi (1957), by Castagnoli, Franzinetti and Manfredini (1957), by Abashian *et al.* (1957) and by others. The actual degree of asymmetry observed depends on the material in which the muon stops, owing to depolarization of the muon to an extent depending on the structure of the material. In graphite the μ^+ meson appears to suffer rather little depolarization before decay, whereas in diamond the depolarization is almost complete. Studies on this depolarization process will undoubtedly give information of great value to solid state physics in the near future.

These asymmetry measurements have been made in two ways. In emulsions and bubble chambers, $\pi^+\rightarrow\mu^+\rightarrow e^+$ is observed and the electron direction correlated with the direction of emission of the muon. Taking together the data from all laboratories, this angular correlation has the form (Kaplon 1957)

$$1-0{\cdot}15(\pm 0{\cdot}01)\cos\theta \qquad \ldots\ldots(\mathrm{A6})$$

in emulsion, averaged over the complete electron spectrum. Muon beams obtained from cyclotrons have also been extensively used for asymmetry experiments. The muons emerging from the cyclotron with a given direction and velocity have resulted from pion decays occurring in various directions. The net muon polarization, which can only be directed parallel or antiparallel to the beam direction, will therefore be less than complete, being non-zero only because the beams selected consist predominantly of muons which were emitted in forward directions from their parent pions. The asymmetry obtained in the decay of μ^+ mesons from the Chicago beam, for example, brought to rest in emulsion leads to a coefficient $-0{\cdot}10(\pm 0{\cdot}02)$; comparison with (A6) then implies that this beam has $65(\pm 15)\%$ polarization. For quite a number of materials the maximum asymmetry observed for μ^+ decay from this beam was obtained, the asymmetry coefficient being $-0{\cdot}24(\pm 0{\cdot}01)$. If the depolarization of the muon is assumed to be zero in these materials, then these results correspond to an asymmetry coefficient $-0{\cdot}36(\pm 0{\cdot}08)$ in the decay of fully polarized muons.

It was pointed out by Lee and Yang (1957) that, assuming the two-component neutrino theory, the neutral particles in muon decay must be neutrino and antineutrino,

$$\mu^+ \rightarrow e^+ + \nu + \bar{\nu}, \qquad \ldots\ldots(A7)$$

for consistency with the experimental results. The alternative possibilities $e^+ + 2\nu$ or $e^+ + 2\bar{\nu}$ would each lead to an electron spectrum characterized by a Michel parameter $\rho = 0$, whereas the best experimental values available at present are the values $\rho = 0{\cdot}67 \pm 0{\cdot}05$ (Rosenson 1957) and $\rho = 0{\cdot}68 \pm 0{\cdot}01$ (Crowe 1957). With (A7), there are only two couplings, V and A, possible. The electron spectrum and angular distribution then takes the form

$$x^2\left[3 - 2x + \frac{2\mathscr{R}(f_A f_V^*)}{|f_A|^2 + |f_V|^2} P(1-2x)\cos\theta\right] dx\, d(\cos\theta), \qquad \ldots\ldots(A8)$$

where $E_e = x(E_e)_{\max}$ and P denotes the polarization of the muon. The Michel parameter for the unpolarized spectrum (A8) is $\rho = 0{\cdot}75$, as given first by Salam (1957). The energy dependence of the asymmetry coefficient of (A8) agrees well with that found (above about 20 MeV) by the Chicago and Columbia groups. The greatest value for the asymmetry coefficient (averaged over the full spectrum) consistent with (A8) is 0·33, obtained for $f_A = f_V =$ real.

At the present time none of the data on the π–μ–e sequence are really inconsistent with the predictions of the two-component neutrino theory or with CP invariance (which would imply f_A and f_V real). On the other hand, the data do not necessitate assuming the two-component theory, and the theoretical predictions are rather insensitive to violation of CP invariance. However, these new phenomena clearly give us a new means to explore the properties of the muon and of its decay and capture interactions. Observations have already been made on its magnetic moment (Garwin, Lederman and Weinrich 1957), on the longitudinal polarization of its decay electrons (cf. Kinoshita and Sirlin 1957), on the nuclear capture of polarized μ^- mesons (cf. Jackson, Treiman and Wyld 1957 b) and on other processes, but the discussion of such experiments is not our present purpose.

Implications for K-particle and Hyperon Phenomena

With the failure of parity conservation and charge conjugation invariance demonstrated for a number of weak interactions, there is no longer any good reason to expect these invariance properties to hold for the decay processes of K-particles and hyperons. The simplest and most likely interpretation of the τ–θ puzzle discussed in § 7 is now that there exists just one variety of K-meson, with many alternative decay modes; the fact that the two pions from the $K_{\pi 2}$ mode have opposite parity from the three pions from the τ mode is now simply an explicit demonstration of the parity non-conservation property of the K-particle decay interactions. This is especially the case in view of the complete absence of any of the phenomena characteristic (cf. § 7.3) of the presence of parity doublets.

Consequently the situation concerning the strong interactions of K-particles and hyperons is now greatly simplified. The parity of the K-particle, although definite, is now a matter of convention since it cannot be defined from the decay processes (which do not conserve parity) whilst the strong interactions always involve pairs of strange particles. What is physically significant is the relative parity between the various hyperons and the K-particle, for it is this which is important for the qualitative characteristics of their mutual interactions (assuming their spins known). It is appropriate to remark here that the attractive proposals of Schwinger, discussed in § 7.4 and § 9.1, are now excluded, since these depended on the assumption of parity doublets.

However, it should be pointed out here that the only manifestation of parity non-conservation yet established for the strange particle decay interactions is this competition of τ and θ decay from the same parent K-meson. No explicit evidence such as the observation of interference between final states of opposite parity resulting from strange particle decay has yet been found for any pion decay modes. Of course, the $K_{\mu2} \rightarrow \mu \rightarrow e$ decay sequence has a direct analogy with the π–μ–e sequence, and a similar backward–forward asymmetry may be sought in the beta-decay of the muon from $K_{\mu2}$ decay. However, the existence of such an asymmetry may well be related simply with the properties of the neutrino and would have no necessary implications for the pion decay modes of strange particles.

In $K_{\pi2}$ decay, the final 2π state always has a definite parity $(-1)^j$, given by its total angular momentum j. Hence no backward–forward asymmetry can ever exist in $K_{\pi2}$ decay, whatever the K-particle spin. Similarly the final 3π state from the τ mode necessarily has even parity if the K-particle spin is zero. However, for non-zero K-spin, for example with spin 2 (which is not yet finally excluded), the three final pions can exist in configurations of both parities for the given total angular momentum. If no account is taken of the initial K-spin direction, then still no backward–forward asymmetry can result, since there is no interference between these two final states—this is obvious from the fact that no pseudoscalar can be constructed from the only two vectors $\mathbf{p}$ and $\mathbf{q}$ in the situation. Even with a polarization for the initial K-mesons, the interference terms may be shown to vanish if time reversal invariance holds in the decay process and if the phases introduced by pion–pion interactions between the outgoing pions may be neglected (as seems rather probable). However, in any case, there is at present no reason to suspect that K-mesons may have a non-zero spin ; all the available data are quite consistent with spin zero, and the main motivation for considering higher spin values, the τ–θ puzzle, no longer exists.

It is interesting to note that in $K_{\mu3}$ and K_{e3} decay the final muon or electron will generally be polarized longitudinally. Measurement of this polarization as function of the particle energy will provide important information bearing on the nature of these decay interactions, especially if the two-component theory of the neutrino is accepted. These situations have been discussed by Werle (1957) and by McDowell (private communication 1957).

According to a theorem of Lee, Oehme and Yang (1957), the observation of the relationship between K^+ and K^- decay properties will shed no light on the invariance of decay interactions under C. This theorem depends only on the reality of the interaction energy and the Lorentz invariant, local character of the underlying theory, and shows that, to lowest order in the weak decay interactions (a very adequate approximation since the coupling parameters effective are of order 10^{-12}), the branching ratios and decay modes for the K^--particle are necessarily the same as for the K^+-particle, irrespective of the validity of invariance under C.

Even the K_1^0, K_2^0 situation for neutral K-particles is not seriously affected if invariance under C does not hold for decay interactions. It was already remarked in § 6.2 that the operation CP necessarily has the value $+1$ for two pions (with total angular momentum $j, C=(-1)^j$ and $P=(-1)^j$) so that only one of the two states (5.8) can decay to two pions if CP is conserved (since these initial states K_1^0, K_2^0 are also eigenstates of CP). In other words, invariance under time reversal is sufficient to ensure that there is a short-lived K_1^0 state decaying to two pions and a long-lived K_2^0 state for which 2π decay is forbidden. In this situation (assuming spin 0), the τ^0 configurations analogous to those occurring for τ^+ decay (which were referred to as 'τ_1^0 configurations' in § 6.3) can result only from the long-lived K_2^0-decay, since for these configurations $C=+1, P=-1$, therefore $CP=-1$; a 3π-decay mode is possible also from the K_1^0 state but only through the highly improbable 'τ_2^0' configurations of § 6.3. Lande, Lederman and Chinowsky (1957) have recently found several decay events in the long-lived K_2^0 beam which represent decay to $(\pi^+ + \pi^- +$ neutral particle) and which are quite consistent with τ^0 decay (note that decay to $(\pi^+ + \pi^- + \gamma)$ is very improbable for a spin zero particle, cf. Dalitz 1955 a).

Further, Lee, Oehme and Yang (1957) have pointed out that even if time reversal T fails, there will still be two mixed $K^0, \bar{K}^0$ states with different lifetimes, but such that the 2π decay mode is permitted for both lifetimes. In this case, a charge asymmetry between (for example) the $(e^+ + \pi^- + \nu)$ and $(e^- + \pi^+ + \bar{\nu})$ modes is to be expected in the long-lived K_2^0 beam. However, as pointed out by Wyld and Treiman (1957), the large lifetime ratio observed between the two neutral K-particle states would allow only a very small charge asymmetry and only very infrequent 2π-decay in the long-lived decay modes. The study of the $K^0, \bar{K}^0$ decay situation provides a very insensitive test for invariance under time-reversal.

In seeking direct evidence of parity non-conservation in the strange particle decay interactions, the hyperon decay modes appear more promising. Of course, the K-particle and the hyperon interactions necessarily have the same invariance (or non-invariance) properties since they are linked by the strong, 'associated-production' interactions for which all these invariance properties hold valid. Assuming hyperon spin 1/2, the pion emission in decay may be into the s-wave or p-wave channel, since parity is not conserved. For a polarized hyperon, the angular distribution of these pions is then

$$\left\{1+\frac{2\mathscr{R}(M_s^* M_p)}{|M_s|^2+|M_p|^2}\langle\boldsymbol{\sigma}\rangle\cdot\hat{\mathbf{p}}_\pi\right\}, \qquad \ldots\ldots(A9)$$

where M_s, M_p denote the s- and p-wave matrix elements. In general a forward-backward asymmetry relative to the hyperon spin is to be expected in hyperon decays. However, its absence could be due to a dominance of s-wave emission over p-wave emission, $M_s \gg M_p$ (or vice versa), or simply to the smallness of the hyperon polarizations in the production processes considered. In the $K^- + P$ reaction for K^--particles at rest, parity is necessarily conserved since the reaction depends on strong interactions, and the resulting hyperons are produced with zero polarization, since no production plane is defined. In the $\pi^- + P$ reactions studied for pions from 0·87 to 1·3 GeV, the Λ-hyperons (or Σ-hyperons) produced could have substantial polarizations despite their strongly backward-peaked (or forward-peaked) angular distribution (this would be especially true if attention were confined to angles θ_K in the region of about 45° to 90° (or 90° to 135° for the Σ-processes)). No up–down asymmetries have been established in some hundreds of Λ and Σ decays examined following the $\pi^- + P$ reaction (however, the standard deviation on the asymmetry coefficient is still only $\pm 0{\cdot}10$). Intuitively it is natural to expect p-wave emission to be less probable than s-wave emission, as a result of the centrifugal barrier faced by outgoing p-wave pions. This would suggest a ratio (equation (4.12)) of order

$$M_p/M_s \sim (p_\pi R)/3 \sim 0{\cdot}05 \text{ to } 0{\cdot}1,$$

which would allow at most 0·2 for the asymmetry coefficient, assuming fully polarized Λ or Σ hyperons. For these reasons, the absence of this asymmetry in the Λ and Σ decays observed is not at all inconsistent with parity non-conservation in strange particle decays.

A further indication of the ratio M_p/M_s for Λ^0 decay can in principle be obtained from the Ruderman–Karplus argument (cf. § 6.2). This must be modified to take into account the emission of both s- and p-wave pions from the Λ^0-hyperon in both the mesonic decay and the non-mesonic capture processes. However, since the final states reached are of opposite parity they do not interfere in the total decay probabilities, either mesonic or non-mesonic. For spin 1/2, the expected internal conversion coefficient R is therefore simply the weighted average

$$R = (R_0 + xR_1)/(1+x), \qquad \ldots\ldots(\text{A10})$$

where x is the ratio $(M_p/M_s)^2$ in free Λ^0 decay. The fact that the observed ratio R for helium hypernuclei (given as 1·5 by Schneps, Fry and Swami (1957)) is close to the computed value R_0 again suggests that s-wave emission is strongly dominant in Λ^0 decay, the corresponding ratio M_p/M_s being of the order of 0·2.

Finally, the discussion of ${}_{\Lambda}H^4$ decay in § 8.2 needs to be modified in view of parity non-conservation in Λ^0 decay. If the spin of ${}_{\Lambda}H^4$ is j, then the pion emitted in the decay

$${}_{\Lambda}H^4 \rightarrow He_4 + \pi^- \qquad \ldots\ldots(\text{A11})$$

necessarily has angular momentum j. Since the Λ^0-particle in ${}_{\Lambda}H^4$ is in an s-state relative to the H^3 core, pion emission into the s-wave or p-wave channel in the Λ^0 decay can give only an s-wave or p-wave pion respectively in the mode (A10). This restriction does not hold for other ${}_{\Lambda}H^4$ decay modes, such as $H_3 + P + \pi^-$, where the pion can be emitted through either s- or p-wave channel. Thus if the s-wave channel is strongly dominant, as suggested by the arguments above, then, for a ${}_{\Lambda}H^4$ spin $j = 1$, this decay mode $(He_4 + \pi^-)$ would be very rare since the pion would have orbital angular momentum $l_\pi = 1$, which would be possible only through the weak p-wave channel of Λ^0 decay. In fact, as shown in table 7, the ($He_4+\pi^-$) mode accounts for about 60% of the observed ${}_{\Lambda}H^4$ decays and we conclude that $j=0$ for the ${}_{\Lambda}H^4$ hypernucleus. This requires that the Λ^0–nucleon interaction is strongest in the singlet state, with the consequences shown in table 8. This conclusion is compatible with the calculations of the Λ^0–nucleon force by Lichtenberg and Ross (private communication 1957) and by Dallaporta and Ferrara (1957) if these forces are predominantly due to the exchange of two pions. With this conclusion, the orbital angular momenta in the initial and final states of the reactions

$$K^- + He_4 \begin{cases} \rightarrow {}_{\Lambda}H^4 + \pi^0, & \ldots\ldots(A12) \\ \rightarrow {}_{\Lambda}He^4 + \pi^-, & \ldots\ldots(A13) \end{cases}$$

are necessarily the same, so that these reactions can satisfy parity conservation only if the K-particle has opposite parity to the Λ^0-particle (corresponding to a pseudoscalar ($\mathcal{N}\Lambda K$) interaction). These reactions are completely forbidden if the K-particle and Λ^0-hyperon have the same parity (and therefore scalar interaction), and for this reason the study of K^--particle interactions in helium will undoubtedly prove to be of the greatest importance.

REFERENCES

ABASHIAN, A., ADAIR, R. K., COOL, R., ERWIN, A., KOPP, J., LEIPUNER, L., MORRIS, T. W., RAHM, D. C., RAU, R. R., THORNDIKE, A. M., WHITTEMORE, W. L., and WILLIS, W. J., 1957, *Phys. Rev.*, **105**, 1927.

ADAIR, R. K., 1955, *Phys. Rev.*, **100**, 1540.

ALVAREZ, L. W., 1956, *Proceedings of the Sixth Rochester Conference*, § 5 (New York : Interscience). (Report of work by MOYER, OSHER, and PARKER.)

ALVAREZ, L. W., CRAWFORD, F. S., GOOD, M. L., and STEVENSON, M. L., 1956 a, *Phys. Rev.*, **101**, 503 ; 1956 b, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. V, p. 28 ; 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

AMALDI, E., 1956, *Proceedings of the* 1955 *Pisa Conference* (*Nuovo Cim.* Suppl., **4**, 179).

AMMAR, R., FRIEDMAN, J. I., LEVI-SETTI, R., and TELEGDI, V. L., 1957, *Nuovo Cim.*, in the press.

ANDERSON, F., LAWLOR, G., and NEVIN, T. E., 1955 a, *Nuovo Cim.*, **2**, 605 ; 1955 b, *Ibid.*, **2**, 608.

D'ANDLAU, C., ARMENTEROS, R., ASTIER, A., DE STAEBLER, H. C., GREGORY, B. P., LEPRINCE-RINGUET, L., MULLER, F., PEYROU, C., and TINLOT, J. H., 1956, *Nuovo Cim.*, **4**, 917.

ARMENTEROS, R., BARKER, K. H., BUTLER, C. C., and CACHON, A., 1951, *Phil. Mag.*, **42**, 1113.

BALDO, M., BELLIBONI, G., CECCARELLI, M., GRILLI, M., SECHI, B., VITALE, B., and ZORN, G. T., 1955, *Nuovo Cim.*, **1**, 1180.

BALDO-CEOLIN, M., BONETTI, A., GREENING, W. D. B., LIMENTANI, S., MERLIN, M., and VANDERNAEGE, G., 1957, *Nuovo Cim.*, in the press.

BALDO-CEOLIN, M., DILWORTH, C. C., FRY, W. F., GREENING, W. D. B., HUZITA, H., LIMENTANI, S., and SICHIROLLO, A. E., 1957, *Nuovo Cim.*, in the press.

BANDTEL, K. C., FRANK, W. J., MADEY, R., and MOYER, B. J., 1954, *Phys. Rev.*, **95**, 639.

BARKAS, W. H., HECKMAN, H., and SMITH, F. M., 1956, *Bull. Amer. Phys. Soc.*, **1**, 184.

BARKAS, W. H., BARRETT, P. H., CUER, P., HECKMAN, H., SMITH, F. M., and TICHO, H. K., 1956, *Phys. Rev.*, **102**, 583.

BARSHAY, S., 1956, *Phys. Rev.*, **104**, 853.

BARTLETT, M. S., 1953, *Phil. Mag.*, **44**, 249.

BERESTETSKI, V. B., 1953, *Doklady Akad. Nauk. S.S.S.R.*, **92**, 519.

BETHE, H. A., and DE HOFFMANN, F., 1955, *Mesons and Fields*, Vol. II (New York: Row, Peterson).

BHOWMIK, B. D., EVANS, D., VAN HEERDEN, I. J., and PROWSE, D. J., 1956, *Nuovo Cim.*, **3**, 574; and private communication (1956).

BIRGE, R. W., PERKINS, D. H., PETERSON, J. R., STORK, D. H., and WHITEHEAD, M. N., 1956, *Nuovo Cim.*, **4**, 834.

BISWAS, N. N., CECCARELLI-FABBRICHESI, L., CECCARELLI, M., CRESTI, M., GOTTSTEIN, K., VARSHNEYA, N. C., and WALOSCHEK, P., 1956, *Nuovo Cim.*, **3**, 1481.

BISWAS, N. N., CECCARELLI-FABBRICHESI, L., CECCARELLI, M., GOTTSTEIN, K., VARSHNEYA, N. C., and WALOSCHEK, P., 1956 a, *Nuovo Cim.*, **3**, 825; 1956 b, *Ibid.*, **4**, 631; 1957, *Ibid.*, **5**, 123.

BLATT, J. M., and WEISSKOPF, V. F., 1952, *Theoretical Nuclear Physics* (London: Chapman and Hall).

BLEWETT, J. P., 1956, *Rep. Progr. Phys.*, **19**, 37 (London: Physical Society).

BLOCK, M. M., and KING, D. T., 1955, *Phys. Rev.*, **97**, 1415.

BLOCK, M. M., HARTH, E. M., BLEVINS, M. E., and SLAUGHTER, G. G., 1956, *Nuovo Cim.*, **4**, 46.

BLUDMAN, S. A., 1956, *Phys. Rev.*, **102**, 1420.

BLUMENFELD, H., BOOTH, E. T., LEDERMAN, L. M., and CHINOWSKY, W., 1956, *Phys. Rev.*, **102**, 1184.

BONETTI, A., LEVI-SETTI, R., PANETTI, M., and TOMASINI, G., 1953, *Nuovo Cim.*, **10**, 1736.

BONETTI, A., LEVI-SETTI, R., PANETTI, M., SCARSI, L., and TOMASINI, G., 1954, *Nuovo Cim.*, **11**, 210, 330.

BOUCHIAT, C., and MICHEL, L., 1957, *Phys. Rev.*, **106**, 170.

BRENE, N., HANSEN, K. H., HOOPER, J. E., and SCHARFF, M., 1956, *Nuovo Cim.*, **4**, 1059.

BRISBOUT, F., FRIEDLANDER, M. W., and IREDALE, P., 1956, *Nuovo Cim.*, **4**, 948.

BROWN, R. H., CAMERINI, U., FOWLER, P. H., MUIRHEAD, H., POWELL, C. F., and RITSON, D. M., 1949, *Nature, Lond.*, **163**, 82.

BRUECKNER, K., SERBER, R., and WATSON, K. M., 1951, *Phys. Rev.*, **81**, 575.

BUDDE, R., CHRETIEN, M., LEITNER, J., SAMIOS, N. P., SCHWARTZ, M., and STEINBERGER, J., 1956, *Phys. Rev.*, **103**, 1827.

BURCHAM, W., 1955, *Progr. Nucl. Phys.*, **4**, 171.

CASE, K. M., 1956, *Phys. Rev.*, **103**, 1449.

CASE, K. M., KARPLUS, R., and YANG, C. N., 1956, *Phys. Rev.*, **101**, 358.

CASSEN, B., and CONDON, E. U., 1936, *Phys. Rev.*, **50**, 846.

CASTAGNOLI, C., CORTINI, G., and FRANZINETTI, C., 1955, *Nuovo Cim.*, **2**, 550.

CASTAGNOLI, C., FRANZINETTI, C., and MANFREDINI, A., 1957, *Nuovo Cim.*, **5**, 684.

CECCARELLI, M., DALLAPORTA, N., GRILLI, M., MERLIN, M., SALANDIN, G., SECHI, B., and LADU, M., 1955, *Nuovo Cim.*, **2**, 542.

CHUPP, W., GOLDHABER, G., GOLDHABER, S., and WEBB, F. H., 1955, *Interactions of Negative K-particles at Rest*, University of California, UCRL-3044.

COCCONI, G., PUPPI, G., QUARENI, G., and STANGHELLINI, A., 1957, *Nuovo Cim.*, **5**, 172.

COESTER, F., 1951, *Phys. Rev.*, **84**, 1259.

COOPER, W. A., FILTHUTH, H., NEWTH, J. A., PETRUCCI, G., SALMERON, R. A., and ZICHICHI, A., 1956, *Nuovo Cim.*, **4**, 1433.

COSTA, G., and DALLAPORTA, N., 1955, *Nuovo Cim.*, **2**, 519.

COURANT, H., 1955, *Phys. Rev.*, **99**, 282.

COWAN, E. W., 1954, *Phys. Rev.*, **94**, 161.

CROWE, K. M., 1957, *Proceedings of the Seventh Rochester Conference* (quoted by L. M. Lederman) (New York: Interscience).

CRUSSARD, J., 1956, *Proceedings of the Sixth Rochester Conference* (New York: Interscience), Chap. V, p. 18.

CRUSSARD, J., FOUCHE, V., HENESSY, J., KAYAS, G., LEPRINCE-RINGUET, L., MORELLET, D., and RENARD, F., 1956 a, *Nuovo Cim.*, **3**, 731; 1956 b, *C.R. Acad. Sci., Paris*, **242**, 1554.

CRUSSARD, J., KAPLON, M. F., KLARMANN, J., and NOON, J. H., 1954, *Phys. Rev.*, **93**, 253.

DALITZ, R. H., 1953, *Phil. Mag.*, **44**, 1068 ; 1954, *Phys. Rev.*, **94**, 1046 ; 1955 a, *Ibid.*, **99**, 915 ; 1955 b, *Ibid.*, **99**, 1475 ; 1956 a, *Nuclear Physics*, **1**, 372 ; 1956 b, *Proc. Phys. Soc.* A, **69**, 527.

DALLAPORTA, N., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. VI, p. 11 ; 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

DALLAPORTA, N., and FERRARA, F., 1957, *Nuovo Cim.*, **5**, 111.

DANIELS, R. R., and PAL, Y., 1954, *Proc. Ind. Acad. Sci.*, **40**, 114.

DANYSZ, M., and PNIEWSKI, J., 1953, *Phil. Mag.*, **44**, 348.

DEBENEDETTI, A., GARELLI, C. M., TALLONE, L., and VIGONE, M., 1954 a, *Nuovo Cim.*, **12**, 466 ; 1954 b, *Ibid.*, **12**, 952.

DERRICK, G. H., 1956, *Nuovo Cim.*, **4**, 565.

EISENBERG, Y., 1954, *Phys. Rev.*, **96**, 541 ; 1956, *Proceedings of the* 1955 *Pisa Conference* (*Nuovo Cim.* Suppl., **4**, 484).

EISENBERG, Y., LOMON, E., and ROSENDORFF, S., 1956, *Nuovo Cim.*, **4**, 610.

EKSPONG, A. G., and GOLDHABER, G., 1956, *Phys. Rev.*, **102**, 1187.

D'ESPAGNAT, B., and PRENTKI, J., 1956 a, *Nuclear Physics*, **1**, 33 ; 1956 b, *Nuovo Cim.*, **3**, 1045.

FABRI, E., 1954, *Nuovo Cim.*, **11**, 479.

FELD, B. T., ODIAN, A. C., RITSON, D. M., and WATTENBERG, A., 1955, *Phys. Rev.*, **100**, 1539.

FELDMAN, D., 1956, *Phys. Rev.*, **103**, 254.

FERMI, E., 1951, *Elementary Particles* (Newhaven : Yale University Press).

FITCH, V., and MOTLEY, R., 1956, *Phys. Rev.*, **101**, 496.

FOURNET, D. M., and WIDGOFF, M., 1956, *Phys. Rev.*, **102**, 929.

FOWLER, P. H., and HANSEN, K. H., 1956, *Nuovo Cim.*, **4**, 158.

FOWLER, T. K., 1956, *Phys. Rev.*, **102**, 844.

FOWLER, W. B., SHUTT, R. P., THORNDIKE, A. M., and WHITTEMORE, W. L., 1953, *Phys. Rev.*, **91**, 1287 ; 1954, *Ibid.*, **93**, 861 ; 1955, *Ibid.*, **98**, 121.

FRAUENFELDER, H., BOBONE, R., VON GOELER, E., LEVINE, N., LEWIS, H. R., PEACOCK, R. N., ROSSI, A., and DE PASQUALI, G., 1957, *Phys. Rev.*, **106**, 386.

FRIEDLANDER, M. W., 1957, *Nuovo Cim.*, **5**, 283.

FRIEDLANDER, M. W., FUJIMOTO, Y., KEEFE, D., and MENON, M. G. K., 1955, *Nuovo Cim.*, **2**, 90.

FRIEDLANDER, M. W., KEEFE, D., and MENON, M. G. K., 1955, *Nuovo Cim.*, **2**, 663.

FRIEDLANDER, M. W., KEEFE, D., MENON, M. G. K., and MERLIN, M., 1954, *Phil. Mag.*, **45**, 533.

FRIEDLANDER, M. W., KEEFE, D., MENON, M. G. K., and VAN ROSSUM, L., 1954, *Phil. Mag.*, **45**, 1043.

FRIEDMAN, J. I., and TELEGDI, V. L., 1957, *Phys. Rev.*, **105**, 1681.

FRY, W. F., SCHNEPS, J., SNOW, G., and SWAMI, M. S., 1955 a, *Phys. Rev.*, **100**, 950 ; 1955 b, *Ibid.*, **100**, 1448 ; 1956, *Ibid.*, **103**, 226.

FRY, W. F., SCHNEPS, J., and SWAMI, M. S., 1955 a, *Phys. Rev.*, **97**, 1189 ; 1955 b, *Nuovo Cim.*, **2**, 346 ; 1955 c, *Phys. Rev.*, **99**, 1561 ; 1956, *Ibid.*, **101**, 1526.

FRY, W. F., and SWAMI, M. S., 1954, *Phys. Rev.*, **96**, 809.

G-STACK COLLABORATION, 1956, *Nuovo Cim.*, **3**, 510.

GARWIN, R. L., LEDERMAN, L. M., and WEINRICH, M., 1957, *Phys. Rev.*, **105**, 1415.

GATTO, R., 1955, *Nuovo Cim.*, **2**, 841 ; 1956 a, *Ibid.*, **3**, 318 ; 1956 b, *Ibid.*, **3**, 1142 ; 1956 c, *Ibid.*, **4**, 197 ; 1956 d, *Ibid.*, **4**, 1274 ; 1957, *Ibid.*, **5**, in the press.

GAYTHER, D. B., 1954, *Phil. Mag.*, **45**, 570.

GELL-MANN, M., 1953, *Phys. Rev.*, **92**, 833 ; 1955, *Proceedings of the Fifth Rochester Conference* (New York : Interscience) ; 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. VIII, p. 23.

GELL-MANN, M., and PAIS, A., 1954, *Nuclear and Meson Physics* (Glasgow : Blackie), p. 342 ; 1955, *Phys. Rev.*, **97**, 1387.

GELL-MANN, M., and WATSON, K. M., 1954, *Ann. Rev. Nucl. Sci.*, **4**, 219.

George, E. P., Herz, A. J., Noon, J. H., and Solntseff, N., 1956, *Nuovo Cim.*, **3**, 94.

Gilbert, F. C., Violet, C. E., and White, R. S., 1956 a, *Phys. Rev.*, **103**, 248 ; 1956 b, *Ibid.*, **103**, 1825.

Glaser, D. A., 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Goldhaber, S., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. VI, p. 1.

Goldhaber, M., Grodzins, L., and Sunyar, A. W., 1957, *Phys. Rev.*, **108**, 826.

Goldsack, S., and Lock, W. O., 1956, *Nuovo Cim.*, **3**, 600.

Gregory, B., Lagarrigue, A., Leprince-Ringuet, L., Muller, F., and Peyrou, Ch., 1954, *Nuovo Cim.*, **11**, 292.

Haddock, R. P., 1956, *Nuovo Cim.*, **4**, 240.

Harris, G., Orear, J., and Taylor, S., 1955, *Phys. Rev.*, **100**, 932 ; 1956, *Ibid.*, **101**, 1213.

Haskin, D. M., Bowen, T., and Schein, M., 1956, *Phys. Rev.*, **103**, 1512.

Haskin, D. M., Bowen, T., Glasser, R. G., and Schein, M., 1956, *Phys. Rev.*, **102**, 244.

Haugerud, O., and Sorensen, S. O., 1955, *Phys. Rev.*, **99**, 106.

Heckman, H. H., Smith, F. M., and Barkas, W. H., 1956 a, *Nuovo Cim.*, **3**, 85 ; 1956 b, *Ibid.*, **4**, 51.

Heckman, H. H., Inman, F. W., Mason, C. J., Nickolls, N. A., Smith, F. M., Barkas, W. H., Dudziak, W. F., and Giles, P. C., 1956, *Bull. Amer. Phys. Soc.*, **1**, 386.

Herz, A. J., Hodgson, P. E., and Tennent, R. M., 1953, *Phil. Mag.*, **44**, 85.

Hildebrand, R. H., 1953, *Phys. Rev.*, **89**, 1090.

Hill, R. D., Salant, E. O., Widgoff, M., Osborne, L. S., Pevsner, A., Ritson, D. M., Crussard, J., and Walker, W. D., 1956, *Phys. Rev.*, **101**, 1127.

Hoang, T. F., Kaplon, M. F., and Yekutieli, G., 1956 a, *Phys. Rev.*, **101**, 1834 ; 1956 b, *Ibid.*, **102**, 1185 ; 1957, *Ibid.*, **105**, 278.

Hodson, A. L., Ballam, J., Arnold, W. H., Harris, D. R., Rau, R. R., Reynolds, G. T., and Treiman, S. B., 1954, *Phys. Rev.*, **96**, 1089.

Hornbostel, J., and Salant, E. O., 1956, *Phys. Rev.*, **102**, 502.

Iloff, E. L., Goldhaber, G., Goldhaber, S., Lanutti, J. E., Gilbert, F. C., Violet, C. E., and Widgoff, M., 1956, *Phys. Rev.*, **102**, 927.

Iwao, S., 1955, *Progr. Theor. Phys., Japan*, **13**, 111.

Jackson, J. D., Treiman, S. B., and Wyld, H. W., 1957 a, *Phys. Rev.*, **106**, 517 ; 1957 b, *Ibid.*, in the press.

Jones, P. T., and Knipp, J. K., 1955, *Nuovo Cim.*, **2**, 857.

Kaplon, M. F., 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Kaplon, M. F., Klarmann, J., and Yekutieli, G., 1955, *Phys. Rev.*, **99**, 1528.

Kawaguchi, M., and Nishijima, K., 1956 a, *Progr. Theor. Phys., Japan*, **15**, 180 ; 1956 b, *Ibid.*, **15**, 182.

Kemmer, N., 1938, *Proc. Camb. Phil. Soc.*, **34**, 354.

Kinoshita, T., and Sirlin, A., 1957, *Phys. Rev.*, in the press.

Koshiba, M., 1956, *Nuovo Cim.*, **4**, 357.

Ladu, M., Levi-Setti, R., and Scarsi, L., 1956, *Proceedings of the 1955 Pisa Conference* (*Nuovo Cim.* Suppl., **4**, 621).

Landau, L. D., 1957, *J. Exp. Theor. Phys., U.S.S.R.*, **32**, 405.

Lande, K., Booth, E. T., Impeduglia, J., Ledermann, L. M., and Chinowsky, W., 1956, *Phys. Rev.*, **103**, 1901.

Lande, K., Lederman, L. M., and Chinowsky, W., 1957, *Phys. Rev.*, **105**, 1925.

Lander, R., Fowler, W. B., and Powell, W. M., 1956, *Bull. Amer. Phys. Soc.*, **1**, 385.

Lanutti, J. E., Chupp, W., Goldhaber, G., Goldhaber, S., Helmy, E., Iloff, E. L., Pevsner, A., and Ritson, D. M., 1956, *Phys. Rev.*, **101**, 1617.

Lee, T. D., 1955, *Phys. Rev.*, **99**, 337.

Lee, T. D., Oehme, R., and Yang, C. N., 1957, *Phys. Rev.*, **106**, 340.

Lee, T. D., and Orear, J., 1955, *Phys. Rev.*, **100**, 932.

Lee, T. D., and Yang, C. N., 1957, *Phys. Rev.*, **105**, 1671.

Lee, T. D., and Yang, C. N., 1956 a, *Phys. Rev.*, **102**, 290 ; 1956 b, *Ibid.*, **104**, 254 ; 1956 c, *Ibid.*, **104**, 822 ; 1957, *Ibid.*, **105**, 1671.

Leighton, R., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience).

Levi-Setti, R., Slater, W., and Telegdi, V. L., 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Lichtenberg, D. B., and Ross, M., 1956, *Phys. Rev.*, **103**, 1131.

Luders, R., 1954, *K. Danske Vidensk. Selsk., Mat-fys. Medd.*, **28**, No. 5.

Marshak, R. E., and Sudarshan, E. C. G., 1957, *Nuovo Cim.*, in the press.

McAllister, R. W., and Hofstadter, R., 1956, *Phys. Rev.*, **102**, 851.

Mezzetti, L., and Keuffel, J. W., 1954, *Phys. Rev.*, **95**, 858.

Moyer, B. J., Osher, J., and Parker, S. I., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. V, p. 32.

Nakano, T., and Nishijima, K., 1953, *Progr. Theor. Phys., Japan*, **10**, 581.

Nishijima, K., 1955 a, *Progr. Theor. Phys., Japan*, **13**, 285 ; 1955 b, *Ibid.*, **14**, 527.

O'Ceallaigh, C., 1951, *Phil. Mag.*, **42**, 1032 ; 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Orear, J., Harris, G., and Taylor, S., 1956, *Phys. Rev.*, **102**, 1676.

Osborne, L. S., 1956, *Phys. Rev.*, **102**, 296.

Page, D. I., 1954, *Phil. Mag.*, **45**, 863.

Page, D. I., and Newth, J. A., 1954, *Phil. Mag.*, **45**, 38.

Page, L. A., and Heinberg, M., 1957, *Bull. Amer. Phys. Soc.*, **2**, 172.

Pais, A., 1952, *Phys. Rev.*, **86**, 663 ; 1954, *Proc. Nat. Acad. Sci., Wash.*, **40**, 484.

Pais, A., and Piccioni, O., 1955, *Phys. Rev.*, **100**, 1487.

Pais, A., and Serber, R., 1955, *Phys. Rev.*, **99**, 1551.

Pauli, W., 1955, *Niels Bohr and the Development of Physics* (London : Pergamon), p. 30.

Peyrou, C., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. VI, p. 36.

Plano, R., Samios, N., Schwartz, M., and Steinberger, J., 1957, *Nuovo Cim.*, **5**, 216.

Primakoff, H., 1956, *Nuovo Cim.*, **3**, 1394.

Primakoff, H., and Cheston, W., 1953, *Phys. Rev.*, **92**, 1537.

Purcell, E. M., and Ramsey, N. F., 1950, *Phys. Rev.*, **78**, 807.

Racah, G., 1956, *Nucl. Phys.*, **1**, 301.

Ridgway, S. L., Berley, D., and Collins, G. B., 1956, *Phys. Rev.*, **104**, 513.

Rochester, G. D., and Butler, C. C., 1947, *Nature, Lond.*, **160**, 855 ; 1953, *Rep. Progr. Phys.*, **16**, 364 (London : Physical Society).

Rose, M. E., 1955, *Multiple Fields* (London : Chapman and Hall).

Rosenson, L., 1957, *Phys. Rev.*, in the press.

Rosselet, P., Weill, R., and Gailloud, M., 1956, *Nuovo Cim.*, **3**, 505.

Ruderman, M., and Karplus, R., 1956, *Phys. Rev.*, **102**, 247.

Sachs, R. G., 1955, *Phys. Rev.*, **99**, 1573.

Sachs, R. G., and Treiman, S. B., 1955, *Nuovo Cim.*, **2**, 1331.

Salam, A., 1956, *Nucl. Phys.*, **2**, 173 ; 1957, *Nuovo Cim.*, **5**, 299.

Salam, A., and Polkinghorne, J. C., 1955, *Nuovo Cim.*, **2**, 685.

Schein, M., Haskin, D. M., and Leenov, D., 1955, *Phys. Rev.*, **100**, 1455.

Schluter, R. A., 1954, *Phys. Rev.*, **96**, 734.

Schneps, J., Fry, W. F., and Swami, M. S., 1957, *Phys. Rev.*, **106**, 1062.

Schopper, H., 1957, *Phil. Mag.*, **2**, 576.

Schwartz, M., 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Schwinger, J., 1956, *Phys. Rev.*, **104**, 1164.

Seeman, N., Shapiro, M. M., and Stiller, B., 1955, *Phys. Rev.*, **100**, 1480.

Skjeggestad, O., and Sorensen, S. O., 1956, *Nuovo Cim.*, **3**, 652.

Smith, F. M., Birnbaum, W., and Barkas, W. H., 1953, *Phys. Rev.*, **91**, 765.

Smith, F. M., Heckman, H. H., and Barkas, W. H., 1956, *Composition of a Secondary Particle Beam from the Bevatron*, University of California, UCRL-3289.

Snow, G., 1956, *Phys. Rev.*, **103**, 1111 ; 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Snyder, A., Chang, W. Y., and Gupta, I. C., 1957, *Phys. Rev.*, **106**, 149.

Solmitz, F. T., Alvarez, L. W., Bradner, H., Falk, P., Gow, J. D., Rosenfeld, A. H., and Tripp, R. D., 1956, *Bull. Amer. Phys. Soc.*, **1**, 385.

Sorrels, J. D., Leighton, R. B., and Anderson, C. D., 1955, *Phys. Rev.*, **100**, 1457.

Sudarshan, E. C. G., and Marshak, R. E., 1956, *Phys. Rev.*, **104**, 267.

Takeda, G., 1956, *Phys. Rev.*, **101**, 1547.

Teucher, M., Thirring, W., and Winzeler, H., 1955, *Nuovo Cim.*, **1**, 733.

Thirring, W., 1956, *Nuovo Cim.*, **4**, 666.

Thompson, R. W., 1956, *Summary of Section* A, *Proceedings of the International Congress on Elementary Particles, Pisa*, 1955 (*Nuovo Cim.* Suppl., **4**, 642).

Thompson, R. W., Buskirk, A. V., Etter, L. R., Karzmark, C. J., and Rediker, R. H., 1953, *Phys. Rev.*, **90**, 1122.

Treiman, S. B., 1956 a, *Phys. Rev.*, **101**, 1217 ; 1956 b, *Ibid.*, **104**, 1475.

Treiman, S. B., and Sachs, R. G., 1956, *Phys. Rev.*, **103**, 1545.

Treiman, S. B., and Wyld, H. W., 1955, *Phys. Rev.*, **99**, 1039.

Utiyama, R., and Tobocman, W., 1955, *Phys. Rev.*, **98**, 780.

Van Lint, V. A. J., and Trilling, G. H., 1953, *Phys. Rev.*, **92**, 1089.

Velde, J. C., Brown, J. L., Glaser, D. A., Meyer, D. I., Perl, M. I., Cronin, J. W., De Benedetti, S., and Cool, R., 1957, *Bull. Amer. Phys. Soc.*, **2**, 221.

Walker, W., 1955, *Phys. Rev.*, **98**, 1407.

Walker, W., and Shephard, W., 1956, *Phys. Rev.*, **101**, 1801.

Watson, K. M., 1951, *Phys. Rev.*, **85**, 852.

Watson, K. M., and Stuart, R. N., 1951, *Phys. Rev.*, **82**, 738.

Webb, F. H., Chupp, W. W., Goldhaber, G., and Goldhaber, S., 1956, *Phys. Rev.*, **101**, 1215.

Wentzel, G., 1956 a, *Phys. Rev.*, **101**, 835 ; 1956 b, *Ibid.*, **101**, 1214.

Werle, J., 1957, *Proceedings of the Seventh Rochester Conference* (New York : Interscience).

Weyl, H., 1929, *J. Phys.*, **56**, 330.

Whitehead, M. N., Stork, D. H., Peterson, J. R., Perkins, D. H., and Birge, R. W., 1956, *Preliminary Report on Masses, Decay Modes, Abundances and Energy Spectra of K Mesons*, University of California, UCRL-3295.

Widgoff, M., Shapiro, A. M., Schluter, R., Ritson, D. M., Pevsner, A., and Henri, V., 1956, *Phys. Rev.*, **104**, 811.

Wigner, E. P., 1932, *Nachr. Ges. Wiss. Göttingen*, **31**, 546.

Wilkinson, D. H., 1953, *Nature, Lond.*, **172**, 576.

Wolfenstein, L., 1949, *Phys. Rev.*, **75**, 1664.

Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D., and Hudson, R. P., 1957, *Phys. Rev.*, **105**, 1413.

Wyld, H. W., and Treiman, S. B., 1957, *Phys. Rev.*, **106**, 169.

Yang, C. N., 1956, *Proceedings of the Sixth Rochester Conference* (New York : Interscience), Chap. VIII, p. 1.

Yekutieli, G., Kaplon, M. F., and Hoang, T. F., 1956, *Phys. Rev.*, **101**, 506.

York, C. M., Leighton, R. B., and Bjornerund, E. K., 1953, *Phys. Rev.*, **90**, 167.

MICROWAVE PROPERTIES OF SOLIDS

By D. M. S. BAGGULEY and J. OWEN
Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	305
§ 2. Defects in crystals	306
2.1. Double resonance method	306
2.2. Donors and acceptors in semiconductors	308
2.3. Colour centres in alkali halides	309
2.4. Defects produced by irradiation	313
2.5. Transition group ion impurities	319
§ 3. Conductors	319
3.1. Spin resonance from conduction electrons	319
3.2. Dilute alloys of Mn in Cu	324
3.3. Superconducting metals	325
3.4. Anomalous skin effect	329
§ 4. Electrical absorption in semiconductors	331
4.1. General discussion	331
4.2. Experimental techniques	333
4.3. Cyclotron resonance	334
4.4. Plasma effects	337
4.5. Metals	338
4.6. Non-resonant absorption in semiconductors	339
§ 5. Spin resonance from pairs of coupled atoms	340
§ 6. General discussion of extensive coupled systems	345
§ 7. Ferromagnetic metals	349
7.1. Experimental principles	349
7.2. Spectroscopic splitting factors	351
7.3. Anisotropy constants	353
7.4. Line shapes and relaxation terms	354
§ 8. Ferrite materials	355
8.1. General discussion	355
8.2. Magnetic resonance	356
8.3. Relaxation effects	358
8.4. Faraday rotation	359
8.5. Initial permeability	360
§ 9. Antiferromagnetic resonance	364
9.1. Summary of experiments	365
9.2. Theory and comparison with experiment	366
Acknowledgments	372
References	373

Abstract. This report discusses the application of electron magnetic resonance techniques to defect and impurity sites in solids, conduction electrons in metals, ferromagnetic, ferrimagnetic and antiferromagnetic materials. It also reviews the recent experiments on cyclotron resonance in semiconductors together with a brief account of some experiments involving the non-resonant absorption of microwaves by conductors at low temperatures and by ferrites.

§ 1. INTRODUCTION

THE use of centimetre waves for the investigation of the solid state is now well established. This is a large field which is expanding rapidly at the present time, and includes electron magnetic resonance, cyclotron resonance and a number of non-resonant phenomena. A particular aspect of electron resonance, namely paramagnetic resonance in transition group salts, has already been reviewed in previous *Progress Reports* (Bleaney and Stevens 1953, Bowers and Owen 1955). In the present report we shall only discuss the applications of electron magnetic resonance techniques to defect and impurity sites in solids, conduction electrons in metals, ferromagnetic, ferrimagnetic and anti-ferromagnetic materials. We also review the recent experiments on cyclotron resonance in semiconductors together with a brief discussion of some experiments involving the non-resonant absorption of microwaves by conductors at low temperatures and by ferrites. We have not however included any account of the measurements of dielectric constants, magnetic resonance from free radicals, the Overhauser (1953) method for nuclear polarization, nor the solid state maser.

In the case of electron magnetic resonance the energy levels of a system possessing a resultant spin angular momentum may be resolved by the application of a magnetic field H. This means that each spin system will have a characteristic set of energy states determined by the field H and an assembly of such systems will be distributed among these states in accordance with the Maxwell–Boltzmann distribution law. The separation of the energy states is often a linear function of H for the magnetic fields used in microwave spectroscopy and transitions may be induced between them by a quantum of radiation ($\hbar\omega$) when $\hbar\omega = g\beta H$. In this relation g is a numerical parameter called the spectroscopic splitting factor, β is the Bohr magneton and it is usually necessary for the radiation to have a component of high-frequency magnetic field perpendicular to the applied field H. When the assembly is in thermal equilibrium the induced transitions give rise to a net power absorption and in many practical cases this occurs at a field and frequency related by $g \sim 2{\cdot}0$. In the case of an ideal free electron spin $g_s = 2{\cdot}0023$ so that for 3 cm microwave radiation $H \sim 3000$ oersteds.

The phenomenon of cyclotron resonance on the other hand arises from the interaction between the high-frequency electric field and the charge carriers in a semiconductor. In a simple case a free charge q having a mass m will spiral around the magnetic field direction at an angular frequency of $\omega = qH/mc$. As in the case of the cyclotron particle accelerator, the charge may absorb energy from a radiation field having this angular frequency ω. For the special case of a free electron ($g_s = 2{\cdot}0023$) the spin resonance and the cyclotron resonance conditions are practically identical and so the two absorption lines coincide. In solids, however, the translational motion of the electrons is strongly influenced by interactions with the crystal lattice and this effect is frequently represented by a modified electron mass m^* (the 'effective mass') which may be only 1/10 of the free electron mass m. The cyclotron resonance then occurs at a field and frequency given by $\omega = eH/m^*c$ so that the spin and cyclotron lines are well separated. At 3 cm wavelengths the cyclotron lines occur in fields ranging from about 300 to 1500 oersteds.

Reviews which cover the recent literature on nuclear and electronic resonance in solids, liquids and gases have been given by Hutchison (1956) and Wertz (1955). This field is also dealt with in the book by Ingram (1955), which includes as well a useful account of experimental techniques.

§ 2. DEFECTS IN CRYSTALS

Any imperfection in a crystal which has one or more unpaired electrons associated with it gives rise, in principle, to a paramagnetic resonance spectrum. Types of defect which have been investigated by the resonance technique include (1) donors and acceptors in semiconductors, (2) colour centres in alkali halides, (3) defects produced by irradiation, (4) transition group ion impurities, (5) free radicals. These classifications overlap to some extent ; for example, some colour centres can be produced by irradiation, and also many radiation defects involving an unpaired electron can be classed as free radicals.

The two main problems in the investigation of defect spectra are first to identify the defect, and then to get detailed information about its electronic state. Features of the method which help towards these ends include the following : (*a*) Different types of defect in the same crystal in general give distinguishable microwave spectra. (*b*) The intensity of a particular spectrum indicates the number of defects of that type. (*c*) The shift of the spectroscopic splitting factor from the free spin value, $\Delta g = g - 2{\cdot}0023$, usually indicates the strength of spin–orbit interaction effects for the unpaired electron, and Δg is normally negative for electrons and positive for holes. (*d*) Any nucleus of spin I associated with the defect is labelled by the $2I+1$ fold hyperfine splitting in the spectrum ; the magnitude of the splitting indicates the time spent by the unpaired electron near this nucleus. This is perhaps the most useful feature of the method. Under certain conditions the resolving power for hyperfine structure can be enormously increased by the application of ‘ double resonance ’ techniques (Feher 1956), as described below. (*e*) The behaviour of the spectrum as the crystal is rotated in the external magnetic field gives information about the symmetry properties of the defect. (*f*) The line width ΔH gives information about magnetic and exchange interactions between defects, and spin–lattice relaxation effects. Under favourable conditions, the width due to unresolved hyperfine structure can be distinguished by the ‘ inhomogeneous broadening ’ (Portis 1953) produced when the absorption line is saturated by the radio-frequency power.

2.1. *Double Resonance Method*

The experimental techniques used in the investigation of electron spin resonance from paramagnetic centres in non-conducting crystals is well established (for a review, see, for example, Bleaney and Stevens 1953), and we will give no details of the general method here. However, we will describe one recent development which is likely to have important applications to paramagnetic resonance problems. This is the double resonance technique as used by Feher (1956) (see also Burgess and Norberg 1955). It is essentially an observation of nuclear magnetic resonance via the electron spin resonance line. The technique is to use sufficient microwave power to saturate the electron line so that its intensity is greatly reduced, and then to enhance this line again by applying a nuclear resonance frequency which corresponds to the hyperfine splitting of the energy levels.

A suitable example, described by Feher, is the application of the method to phosphorus-doped silicon, where (see § 2.2) there is one unpaired electron moving in an orbit round each ^{31}P nucleus, and the electron resonance line is easy to saturate at liquid helium temperatures since the spin lattice relaxation time is many seconds. The energy levels in magnetic field H are given by the Hamiltonian†

$$\mathscr{H} = g\beta \mathbf{H} . \mathbf{S} + A\mathbf{S} . \mathbf{I} - g_N \beta_N \mathbf{H} . \mathbf{I},$$

where $g = 2{\cdot}00$, $S = 1/2$, $I = 1/2$ for ^{31}P and the nuclear g-value $g_N = 2{\cdot}26$, and $A/h = 118$ Mc/s $\rightarrow$ 42 oersteds as given in table 2.1. Feher used a wavelength of about 3·3 cm so that $H \simeq 3000$ oersteds for resonance, and since $g\beta H \gg A$, $g_N \beta_N H$, the energy levels (figure 2.1) are given to a good approximation by

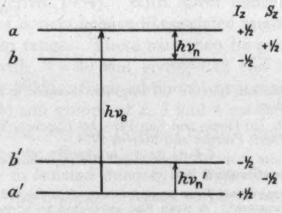


Figure 2.1. Magnetic energy levels for phosphorus in silicon.

$$a, b = \tfrac{1}{2}g\beta H \pm \tfrac{1}{4}A \mp \tfrac{1}{2}g_N \beta_N H$$

$$b', a' = -\tfrac{1}{2}g\beta H \pm \tfrac{1}{4}A \pm \tfrac{1}{2}g_N \beta_N H.$$

The electron spin transition a–$a'(\Delta m_I = 0)$ is now saturated by maintaining the condition $h\nu_e = g\beta H + \frac{1}{2}A$, so that the populations of a and a' are equalized and the absorption intensity goes to zero. A slowly variable nuclear resonance frequency ν_n of order $\frac{1}{2}A/h = 59$ Mc/s is now applied, and when $h\nu_n = \frac{1}{2}A - g_N \beta_N H$, nuclear transitions a–b occur, so that a is depopulated into b, and the electron resonance signal a'–a is restored. Similarly the transition a'–b' with energy $h\nu_n' = \frac{1}{2}A + g_N \beta_N H$ can be measured via the intensity of the electron resonance signal. Feher finds for P in Si that $\nu_n + \nu_n' = A/h = 117{\cdot}6$ Mc/s in agreement with the directly measured value of A (table 2.1), and that

$$\nu_n' - \nu_n = 2g_n \beta_N H/h = 11{\cdot}59 \text{ Mc/s},$$

giving $g_N = 2{\cdot}265$ in good agreement with previous values for the nuclear moment of ^{31}P.

For systems where the electron resonance line can be saturated, this double resonance method clearly has wide application either for a direct measure of nuclear magnetic moments, or for a very accurate measure of the hyperfine spacing of the energy levels, or in some cases for resolving hyperfine structure which in the ordinary direct method is obscured by the width of the electron resonance line. Examples of this last application described by Feher are (i) ^{7}Li hyperfine structure for Li in Si (see § 2.2); effects were also found which he ascribed to the very small hyperfine structure expected from ^{29}Si; (ii) the resolution of hyperfine structure from ^{39}K and ^{35}Cl nuclei in the electron resonance from F-centres in KCl (see § 2.3).

---

† For a discussion of the use of spin-Hamiltonians in paramagnetic resonance work see Bleaney and Stevens (1953) and Bowers and Owen (1955).

Table 2.1. Electron Spin Resonance Data for Donors in Silicon

(1)	(2)	(3)	(4)	(5)
^{7}Li	3/2	3·25	0·3 (a)	1·998 (b)
^{31}P	1/2	1·13	42 (c, a)	2·000 (c)
^{75}As	3/2	1·5	76 (c)	2·0004 (c) ±0·0005
^{121}Sb	5/2	3·36	69 (c)	2·000 (c)
^{123}Sb	7/2	2·55	38 (c)	

(1) Donor nucleus; (2) nuclear spin I; (3) magnetic moment (n.m.); (4) hyperfine structure constant $A/g\beta$ (Oe); (5) g-value.

References: (a) Feher 1956, (b) Honig and Kip 1954, (c) Fletcher, Yager, Pearson, Holden, Read and Merritt 1954, Fletcher, Yager, Pearson and Merritt 1954.

Notes: (i) If there are more than 10^{18} donors/cm^3 the hyperfine structure is narrowed out by exchange and there is only one absorption line (c).

(ii) Authors (c) report no resonance from acceptors in p-type Si, either for 10^{17} boron atoms/cm^3 or for 3×10^{17} aluminium atoms/cm^3. A weak line attributed to surface states has been found in both p- and n-type samples.

(iii) For 10^{16} P atoms/cm^3 in Si the relaxation time $T_1 \simeq 1$ min; for 5×10^{17} P atoms/cm^3, $T_1 < 10^{-5}$ sec; both measurements in temperature range 1·2° to 4°K (Feher and Fletcher 1956).

We now proceed to give a more general account of the resonance spectra from different types of defect.

2.2. *Donors and Acceptors in Semiconductors*

Electron spin resonance has been observed for the interstitial donor Li in Si, and for the substitutional group V donors P, As and Sb in Si (table 2.1). The experiments have been made at low temperatures (~ 4°K) where the semiconductor is effectively an insulator. These impurities have one extra loosely bound unpaired electron compared with the atoms of the host material, and this electron is responsible for the resonance spectrum. The electron can be easily ionized off (by energy of order 0·03 ev) and donated to the conduction band, thus making the silicon n or negative type. Up to the present time, no spin resonance appears to have been found (although looked for) from acceptors such as B, Al, in silicon, nor from donors or acceptors in germanium.

A general theory of these impurity states has been given by Kittel and Mitchell (1954) and by Luttinger and Kohn (1955). The basis of the model is that the loosely bound extra electron (or hole for an acceptor) is moving in a hydrogen-like 1s orbit of large radius R about the impurity nucleus. Using a simple Bohr picture, $R = \epsilon(m/m^*)a$ where $a = \hbar^2/me^2 \simeq 0{\cdot}5$ Å is the usual first Bohr radius for an electron of mass m moving round a nuclear charge $|e|$; ϵ is the dielectric constant, approximately 12 for silicon; m^* is a mean effective mass of order say 0·3m, and represents the fact that the electron is moving in the periodic potential of the Si lattice as well as in the Coulomb potential $-e/\epsilon r$ of the impurity nucleus. The numerical values then give $R \sim 20$ Å, so that the orbit will include several hundred Si atoms.

The experimental spin resonance results give direct evidence for electrons moving in large s-like orbits, and include the following features: (i) Spin–orbit interaction effects are very small, the *g*-shifts being small and, as expected, negative for donors (table 2.1), and the spin–lattice relaxation time being very long. This time is many seconds in the liquid helium temperature range (see references in table 2.1). (ii) There is appreciable overlap, and therefore exchange interaction, between electrons belonging to donors separated by many lattice spaces. For example, if there are more than about 10^{18} donors/cm^3 in Si, this exchange averages out all hyperfine structure leaving a single line of width $\Delta H \simeq 3$ Oe (Fletcher, Yager, Pearson and Merritt 1954). With lower concentrations, say less than 10^{17}/cm^3, the majority of donors behave like isolated atoms, but some have neighbours within the overlap range. There may then be pairs of donors acting as a single magnetic unit with $S = 1$, and giving half the hyperfine spacing of an isolated donor (Slichter 1955). Such additional hyperfine lines have been found by Fletcher *et al.* (1954) and groups of 2, 3 and 4 exchange coupled donors have been detected in a similar way by Feher, Fletcher and Gere (1955). A fuller discussion of these exchange effects is given in § 5.

(iii) Returning now to isolated donors, the hyperfine interaction $A\mathbf{S}\,.\,\mathbf{I}$ of the unpaired electron with the nucleus is found to be of order 1/100 or less of the free atom value, corresponding roughly to the electron being spread over a hundred or more atoms. For example, with ^{7}Li in silicon, the observed spacing is $A/g\beta = 0{\cdot}3$ Oe, whereas the corresponding spacing for the ground state of a free Li atom is $A/g\beta = 144$ Oe. The results are given in table 1 and are in fair agreement (within a factor 2) with the hyperfine structure theory of Kohn and Luttinger (1955).

(iv) In dilute samples, a limit to the line width is set by the small unresolved hyperfine structure from the odd isotopes of Si included in the orbit (^{29}Si, $\mu_I = -0{\cdot}55\beta_N$, 5% natural abundance). For example, Honig and Kip (1954) and Kittel (1954) estimate that for Li in Si the electron orbit includes of order 600 Si atoms, of which the 30 ^{29}Si nuclei give a line width (distance between the half power points) of about 1 Oe. The experimentally observed signal has width 1·5 Oe. It will be noted that this is greater than the ^{7}Li hyperfine spacing $A/g\beta = 0{\cdot}3$ Oe quoted above, so that $2I+1 = 4$ hyperfine lines with adjacent separation 0·3 Oe cannot in fact be directly resolved. They can be resolved, however (Feher 1956), by using the powerful double resonance method as described above.

2.3. *Colour Centres in Alkali Halides*

The colour centres to be discussed are shown in figure 2.2 where the designations are as given by Seitz (1954). An F-centre is a negative ion vacancy (for example, a missing Cl$^-$ ion in KCl) with an extra unpaired electron distributed over the neighbouring ions. A U-centre is a negative ion vacancy containing a negative hydrogen ion H$^-$. A V$_1$-centre is a positive ion vacancy (for example, a missing K$^+$ ion in KCl) with an electron missing from one of the neighbouring ions, or in other words there is an electron hole distributed over the neighbouring ions. More complex centres can be built up if there is more than one vacancy; for example, a V_3-centre consists of two adjacent positive ion vacancies with one electron hole. If this hole is mainly localized over the two nearest neighbours, the centre would be expected to look something like a Cl_2^- molecule ion for the KCl example (see figure 2.2). Electron spin resonance has so far been observed for F-centres, V_3-centres and photo-dissociated U-centres.

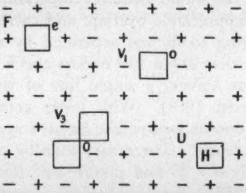


Figure 2.2. Diagram of some colour centres in alkali halide crystal, e.g. K^+Cl^-. The structure is simple cubic. □ = vacancy, e = extra electron, o = missing electron or hole, H^- = negative hydrogen ion. A V_2-centre is similar to a V_3-centre except that there are two holes, and consequently no unpaired electrons.

2.3.1. *F-centres*

The first observation of electron spin resonance was by Hutchison (1949) in neutron irradiated LiF and KCl. Subsequent measurements on F-centres produced by additive colouring or by X-irradiation or electron bombardment at room temperature have been made by other authors, including Schneider and England (1951), Tinkham and Kip (1951), Hutchison and Noble (1952), and particularly by Kip, Kittel, Levy and Portis (1953), whose results are summarized in table 2.2. The main features of the results are as follows:

(i) The *g*-shift is negative as expected for electrons (it would be positive for V-centres) and shows order of magnitude agreement with the spin–orbit interaction contribution to Δg estimated on a molecular orbital model by Kahn and Kittel (1953). This is evidence for the partial p-character of each alkali atom orbit in the LCAO method.

(ii) Contributions to the line width include (*a*) spin–lattice relaxation at room temperature, though this is negligible below 90°K; (*b*) broadening due to dipolar and exchange interaction between centres if the centre concentration is high, for example, in KCl the width is about 50 oersteds if there are 10^{17} centres/cm^3 and about 160 oersteds if there are 10^{19} centres/cm^3; (*c*) there is a residual width due to unresolved hyperfine structure for low concentrations, 10^{16} to 10^{17}/cm^3.

(iii) The magnitude of this hyperfine contribution and the resulting line shape has been carefully analysed by Kip, Kittel, Levy and Portis (1953), who find that their results are consistent with the expected F-centre structure. The results are in table 2.2, and the analysis is further discussed below. At the present time these results are the best direct evidence for the structure of F-centres.†

---

† *Note added in proof*: Better evidence is provided by the recent measurements of hyperfine structure from F-centres in KCl by G. Feher (1957, *Phys. Rev.*, **105**, 1122) and in LiF and NaF by N. W. Lord (1957, *Phys. Rev.*, **105**, 756).

(iv) Further confirmation that the residual line width is due to unresolved hyperfine structure is provided by the work of Portis (1953), who has shown that the absorption lines are broadened inhomogeneously under saturation conditions, and by Feher (1956), who in preliminary measurements has actually been able to resolve this hyperfine structure by using double resonance techniques.

Table 2.2. Electron Spin Resonance Data for F-centres in Alkali Halides (*after* Kip, Kittel, Levy and Portis 1953, and Kittel 1955)

Crystal	g-value	Line width (Oe)		Approximate h.f.s. constants $A'/g\beta$ (Oe)	
		expt	calc	K or Na	^{35}Cl or Br
^{39}K $^{35,(37)}$Cl (natural)	1·995	54	54	7·5, (83)	2·6, (73)
^{41}K $^{35,(37)}$Cl	1·995	36	36	4·2, (45)	2·6, (73)
^{39}K ^{79,81}Br	1·980	146	150	7·5, (83)	~15, (330)
^{23}Na $^{35,(37)}$Cl	1·987	162	190	30, (310)	2·6, (73)

The line width is for low concentrations of F-centres (10^{16} to $10^{17}/\text{cm}^3$), and refers to the separation between points of half maximum absorption. The numbers in brackets are the hyperfine structure constants for the ground state of the corresponding free atom. For explanation of the 'measured' hyperfine structure constants, see text. For similar results on NiF see C. K. Jen and N. W. Lord (1954, *Phys. Rev.*, **96**, 1150).

The molecular orbital theory of F-centre structure has been discussed by a number of authors, including Muto (1949), Inui and Uemura (1950), Kahn and Kittel (1953). For the present purpose we give a simple outline of the treatment by Kip, Kittel, Levy and Portis (1953) and by Kittel (1955), who were particularly concerned with explaining the electron spin resonance line widths and shapes. It should be mentioned that the other possible approach (see, for example, Simpson 1949), in which the unpaired electron is considered to move in an orbit in a homogeneous dielectric medium about the net charge $|e|$ associated with the vacancy (cf. the treatment of donors in Si), can also account for the results provided that allowance is made for the large periodic variation in the crystalline potential over the ionic lattice around the F-centre (Dexter 1954, Krumhansl 1954, Seitz 1954).

Following Kip *et al.*, we now consider the example of an F-centre in KCl. There is a missing Cl^- ion, and as a first approximation the extra unpaired electron is assumed to be distributed only over the six nn K^+ ions. The molecular orbital wave function is

$$\psi = \left(\frac{1}{6}\right)^{\frac{1}{2}} \sum_{i=1}^{6} \psi_i$$

where ψ_i is the orbit on a K^+ ion, and consists of a mixture of 4s and 4p atomic orbits. The hyperfine interaction with each ^{39}K nucleus ($I = 3/2$) is determined mainly by the fraction $\xi^{\frac{1}{2}}$ of 4s orbit on K used in each ψ_i, the contribution from 4p being much smaller. The experimental line widths suggest $\xi \simeq 0{\cdot}6$. In an applied field H, the Hamiltonian for the unpaired electron is then

$$\mathscr{H} = g\beta\mathbf{H}\,.\,\mathbf{S} + \tfrac{1}{6}\xi A\mathbf{S}\,.\sum_{1}^{6}\mathbf{I}_i$$

where $g \simeq 2$, $S = 1/2$, $I_i = 3/2$ for ^{39}K, A = free atom hyperfine structure constant due to 4s electron. The absorption frequencies are given by

$$h\nu = g\beta H + \tfrac{1}{6}\xi A I_z$$

where the total magnetic quantum number $I_z = \sum_1^6 I_{z,i}$, and each $I_{z,i}$ can take values 3/2, 1/2, −1/2, −3/2. There are thus 19 lines corresponding to $I_z = 9, 8, \ldots, -9$, with spacing between adjacent lines $\xi A/6 = A'$ say, and with relative intensities corresponding to the number of ways of forming each I_z value. Thus for

Thus for I_z	= 9	8	7	6	5	4	3	2	1	0
intensity	= 1	6	21	56	120	216	336	456	546	580

and the intensities are symmetrical about $I_z = 0$. This structure is unresolved, but gives an approximately Gaussian line shape (as observed experimentally) and a width between the points of half maximum absorption of about $7A'$ from which it follows that the hyperfine interaction with each ^{39}K nucleus is about 7·5 Oe. Kip *et al.* confirmed this interpretation of the line width by measuring ^{41}KCl where the electron distribution is expected to be the same but the nuclear magnetic moment is smaller $[\mu(39)/\mu(41) = 1{\cdot}78]$. From a comparison of these results they were also able to estimate the contribution to the line width from the hyperfine interaction with the 12 next nearest neighbour ^{35,37}Cl nuclei, and hence to estimate the line widths for KBr and NaCl. These results are summarized in table 2.2.

2.3.2. V-*centres*

Kanzig (1955) has measured the electron spin resonance from V-centres in KCl, NaCl, KBr and LiF. The unpaired electrons were produced by x-irradiation at 90°K, the paramagnetism and the corresponding optical absorption band fading if the temperature was raised much above 120°K. The only detailed results published are for KCl, where Kanzig finds the resonance spectrum to be consistent with that expected from a Cl_2^- ion directed along a (1, 1, 0) axis. Preliminary measurements on NaCl, KBr and LiF suggest similar centres, namely Cl_2^-, Br_2^- and F_2^- ions respectively, directed along a (1, 1, 0) axis. As pointed out by Cohen (1956) this suggests that the centres are of the V_3-type (see figure 2.2 and also remarks above), rather than of the V_1-type† where the symmetry would be cubic and the unpaired electron would be shared between six Cl^- ions as a first approximation.

Kanzig's results on KCl are summarized by the Hamiltonian

$$\mathscr{H} = \beta\mathbf{H}.g.\mathbf{S} + \mathbf{S}.A.(\mathbf{I}_1 + \mathbf{I}_2)$$

where there is an axis of symmetry parallel to any of the six (1, 1, 0) directions (i.e. there are six differently oriented types of magnetic centre), $S = 1/2$, $g_\parallel = 2{\cdot}0023$, $g_\perp = 2{\cdot}024$, $I_1 = I_2 = 3/2$ and $\mu_I = 0{\cdot}82\beta_N$ for ^{35}Cl (the less abundant isotope ^{37}Cl is neglected), and $A_\parallel = 102$ Oe, $A_\perp = 10$ Oe. The absorption frequencies for each type of Cl_2^- ion are then $h\nu = g\beta H + AI_z$ where the total nuclear magnetic quantum number $I_z = I_{z,1} + I_{z,2}$, and $I_{z,1}$ and $I_{z,2}$ can each take the values 3/2, 1/2, −1/2, −3/2. There are thus seven hyperfine lines with adjacent spacing A, and with relative intensities 1 : 2 : 3 : 4 : 3 : 2 : 1. Kanzig finds that the magnitude and anisotropy of A is consistent with the unpaired electron (or rather electron hole) wave function having 30% s and 70% p character. It will be noted that the mean g-value is greater than the free spin value as expected for a hole, and the molecular orbital treatment of a Cl_2^- ion by Inui, Harasarwa and Obata (1956) give theoretical g-values in good agreement with the measured values.

---

† No resonance which can be unambiguously attributed to V_1-centres appears to have been found. Schneider (1954) reports a line of width 90 Oe and $g = 2{\cdot}02$ in x-rayed LiF which he suggests may be from V_1-centres. On the other hand, Portis and Shaltiel (1955) find resonance in x-rayed LiF and LiCl at $g = 2{\cdot}002$, which they attribute to F-centres. The negligible g-shift corresponds to the small spin–orbit coupling for Li.

Recent measurements by Smaller, Delbecq and Yuster (1956) on thallium activated KCl at 80°K after X- or γ-irradiation show centres similar to those found by Kanzig, i.e. Cl_2^- ions along (1, 1, 0) directions. They find that the resonance can be destroyed either by heating to 200°K or by infra-red irradiation.

2.3.3. U-*centres*

Delbecq, Smaller and Yuster (1956) have measured electron spin resonance from photodissociated U-centres in KCl. A U-centre consists of a negative ion vacancy containing a negative hydrogen ion, and can be produced by heating an alkali halide (which already contains F-centres) with the corresponding hydride. Under ultra-violet irradiation, the U-centre is thought to dissociate into an F-centre and a hydrogen atom, $U + h\nu \rightarrow F + H$. Delbecq, Smaller and Yuster found that after irradiation at 80°K the spin resonance spectrum in KCl showed the usual F-centre line plus a hyperfine doublet which they attributed to interstitial hydrogen atoms. The corresponding spectrum from dissociated U-centres prepared using deuterium showed a triplet instead of a doublet. The observed separation between adjacent lines compares with the free atom values as follows :

$^1H, I = 1/2$	$A/g\beta$ (Oe) $= 500 \pm 10$ (in KCl)	506 (free atom),
$^2D, I = 1$	$A/g\beta$ (Oe) $= 78$ (in KCl)	78 (free atom).

The observed g-value was very close to the free spin value, and the line width was about the same as that of the F-centre line, which in the present fairly concentrated samples (2×10^{18} cm^{-3}) was about 68 Oe. No resonance corresponding to H (or D) was found after irradiation at room temperature ; it was presumed that at this temperature the liberated H atoms migrated and combined to form H_2 molecules.

2.4. *Defects produced by Irradiation*

Some examples of irradiation defects involving unpaired electrons which give paramagnetic resonance absorption are shown in table 2.3. This list does not include all of the rapidly growing number of radiation defects which have ever been measured, but does include examples of many of the different types of defect which can be produced. These can be roughly divided into three broad and overlapping classes, some simple examples of which are illustrated in figure 2.3. They are (i) trapping centres, (ii) broken bonds, including free radicals, (iii) lattice damage.

Case (i) usually refers to ionic crystals where there are imperfections (e.g. vacancies or impurities) in the lattice, and the effect of fairly low energy radiation (e.g. X-rays) is to cause a redistribution of electrons so that the trapping centre gains or loses an electron. The initial state can often be restored by thermal energy up to a few hundred degrees Kelvin. Examples are F- and V-centres already discussed above, and quartz with Al impurity ions substituting for Si ions in the SiO_2 lattice.

Case (ii), broken bonds, includes all classes in the sense that any paramagnetic defect produced by irradiation must involve a disturbance of the electron distribution. The term is often used in connection with defects in strongly covalent substances; for example, it is thought that C–C bonds of the polymer chains in some plastics are broken by irradiation (see table 2.3). Some simple possible meanings of 'broken' are shown in figure 2.3. Before irradiation there are two paired electrons in a covalent bond between two atoms, but as a result of irradiation the following processes may occur: (*a*) An electron may be knocked out of the bond. (*b*) One of the electrons may be excited to a higher state, so that the two electrons can have their spins parallel in a triplet (and probably antibonding) state. The state would of course have to be long-lived to give spin resonance absorption. (*c*) The lattice is sufficiently disturbed by the irradiation (e.g. the atoms get pulled apart because of some intermediate process such as (*b*)) to make it energetically unfavourable to form a covalent bond, and each electron is unpaired and can be considered to sit on one of the two atoms. This sort of process is thought to occur in fluon for example, where C–C bonds are broken by x-irradiation. The two radicals produced are very reactive, and their lifetime is limited by their strong tendency to combine with any suitable atom, e.g. oxygen which may diffuse in through the surface of the specimen.

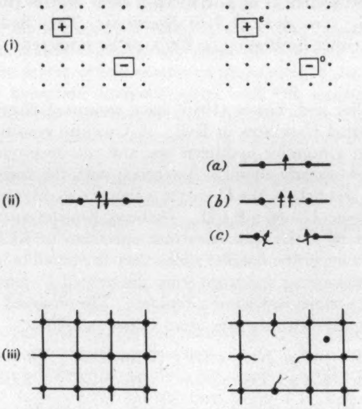


Figure 2.3. Diagram of some simple possible ways of producing unpaired electrons by irradiation. (i) Redistribution of electrons (e) and holes (o) when trapping centres are present, e.g. F-centres, smoky quartz. (ii) Broken bonds, e.g. polymers such as fluon. (iii) Lattice damage, e.g. neutron irradiated diamond.

Table 2.3. **Some Centres produced by Irradiation which show Electron Spin Resonance Absorption**

Substance	Method of irradiation	Centre produced	Hyperfine structure	Ref.
Alkali halides (e.g. KCl)	X at 290°K	F	Unresolved, e.g. from 6 ^{39}K, $I = 3/2$ and 12 ^{35}Cl, $I = 3/2$	See sections 2.2 and 2.3
plus vacancies	X at 90°K	V_3	Resolved, e.g. from 2 ^{35}Cl, $I = 3/2$	
plus U-centres	UV at 77°K	F and atomic H	From ^{1}H, $I = 1/2$	
SiO_2 plus Al impurity	X at 290°K	O^-	Small, from ^{27}Al, $I = 5/2$	a, b, c
$Na(NO_3)$	X at 290°K	NO_2 or NO_3 molecule ?	From ^{14}N, $I = 1$	d
(La, Am, Pm), $Mg_3(NO_3)$, $24H_2O$	α, β, γ at 290°K	NO_2 molecule	From ^{14}N, $I = 1$	e
H_2SO_4 (frozen) D_2SO_4 (frozen)	γ at 77°K	atomic H atomic D	From ^{1}H, $I = 1/2$ From ^{2}D, $I = 1$	f
H_2O (ice) D_2O (heavy ice)	γ at 77°K	H and OH D and OD	From ^{1}H, $I = 1/2$ From ^{2}D, $I = 1$	g
Fluon (or Teflon)	X at 290°K	Broken C–C bond	From 2 ^{19}F, $I = 1/2$	h
			From 7 ^{19}F, $I = 1/2$	i
MgO	X at 290°K, UV at 290°K	O^- ?	None	j k
Diamond	Neutron at 290°K	C vacancies and interstitials, C_2 molecules ?	None	a, b
p-type boron doped silicon	Neutron at 290°K	?	None	l

References: a, Griffiths, Owen and Ward 1955; b, O'Brien and Pryce 1955; c, O'Brien 1955; d, Ard 1955; e, Bleaney, Hayes and Llewellyn 1957; f, Livingston, Zeldes and Taylor 1955, who also discuss resonance from atomic H in other irradiated systems; g, Smaller, Matheson and Yasaitis 1954, Matheson and Smaller 1955; h, Schneider 1955, who also discusses other irradiated plastics (see also Schneider, Day and Stein 1951); i, Ard, Shields and Gordy 1955; j, Wertz and Vivo 1955; k, Griffiths and Orton (unpublished); l, Lark-Horowitz, Schulz-DuBois, Nisenoff and Fan 1955.

Other irradiated substances showing electron spin resonance absorption include: glass, Yasaitis and Smaller 1953, Sands, Townsend and Hood 1954; sugar and wood and plastics, Combrisson and Uebersfeld 1954; various natural proteins (e.g. hair, toe-nail, silk) and crystalline amino and other acids, Gordy, Ard and Shields 1956; frozen hydrogen sulphide and hydrazine chloride, Ard and Gordy 1955.

Case (iii), lattice damage, occurs when bombardment by high-energy heavy particles knocks atoms from their proper place in the lattice, leaving interstitial atoms and vacancies or groups of adjacent vacancies. Examples are diamond and quartz irradiated with neutrons. The original structure can be partly, but usually not entirely, restored by high temperatures of order 1000°c, when the displaced atoms can diffuse back to their proper lattice sites. Interstitial atoms can also be produced via processes (i) and (ii). For example, hydrogen atoms can be liberated from U-centres in KCl by ultra-violet irradiation as already discussed above, or they may be produced by γ-irradiation of frozen sulphuric acid and other substances (table 2.3). A rather intermediate case is that of paramagnetic resonance from NO_2 molecules (table 2.3), where it is assumed that the irradiation removes an oxygen atom from the NO_3 groups of the original crystal. We now give further details of some of the results in table 2.3.

2.4.1. SiO_2 *plus* Al *impurity* (0·01 *to* 0·1% *by weight*).

X-irradiation produces magnetic centres and a smoky coloration which can be removed by heating to 300°c. The spin resonance results of Griffiths, Owen and Ward (1955) show that each centre has one unpaired hole, with $g_\perp = 2{\cdot}00$, $g_\parallel = 2{\cdot}06$, and axis of symmetry approximately parallel to an Si–Si direction. There is a small six line hyperfine structure (attributed to ^{27}Al, $I = 5/2$, $\mu_I = 3{\cdot}64\beta_N$) with average spacing between adjacent lines $A/g\beta \simeq 6$ Oe; this structure shows some anisotropy as do additional lines due to quadrupole effects, and the axis of symmetry is approximately along an Si–O direction which is different from the g-value axis. Griffiths *et al.* suggested the following possible model for the defect: (i) Al^- substitutes for iso-electronic Si (or, using an ionic picture, Al^{3+} for Si^{4+}): it will be noted that this impurity centre is diamagnetic and has a net negative charge with respect to the rest of the crystal; (ii) irradiation removes an electron from this centre; (iii) the remaining unpaired hole is located mainly on an adjacent oxygen ion which then looks something like O^- using an ionic model; (iv) this unpaired spin migrates for a few per cent of the time into an orbit on Al thus giving the small ^{27}Al hyperfine structure. A quantitative treatment of this model is given in the detailed molecular orbital theory of O'Brien (1955) (see also O'Brien and Pryce 1955), who finds reasonable agreement with experiment, particularly with the directional properties of the spectrum. The results illustrate a general rule which helps in recognizing the structure of defects, namely that if the g-value axis of symmetry is different from the hyperfine structure axis, it can usually be assumed that the unpaired electron is not centred about the nucleus responsible for the hyperfine structure; in the present case for example, this was the original reason for supposing that the unpaired spin is located on O rather than on Al. No spin resonance has been found (though looked for) from the electron 'removed' by the irradiation.

2.4.2. *Nitrate salts.*

Ard (1955) finds resonance with a three-line hyperfine structure (due to ^{14}N, $I = 1$) in x-irradiated sodium nitrate, and Bleaney, Hayes and Llewellyn (1957) find similar effects in irradiated lanthanum magnesium nitrate. The latter authors irradiated their specimens with γ-rays from a ^{60}Co source, or by growing crystals containing ^{241}Am (α-emitter) as impurity, or with ^{147}Pm (β-emitter) as impurity. The averaged results can be summarized as follows :

$$\begin{array}{lll}
(a)\ \mathrm{Na(NO_3)} & g = 2{\cdot}000\ \text{(isotropic)} & A/g\beta = 45\ \text{Oe}, \\
(b)\ \mathrm{LaMg_3(NO_3), 24H_2O} & \begin{cases} g_{\parallel} = 1{\cdot}993 \\ g_{\perp} = 2{\cdot}005 \end{cases} & \begin{cases} A_{\parallel}/g\beta = 51\ \text{Oe}, \\ A_{\perp}/g\beta = 64\ \text{Oe}, \end{cases} \\
(c)\ \mathrm{NO_2}\ \text{(free molecule)} & & A/g\beta = 47\ \text{Oe}.
\end{array}$$

It will be noted that the hyperfine structure for the defects (*a*) and (*b*) is fairly close to that for NO_2 molecules in the gaseous state (*c*) (Castle and Beringer 1950). Ard suggests that the defects in $NaNO_3$ may be due to NO_2 or NO_3 molecules, both of which have one unpaired electron. Bleaney *et al.* (1957) suggest that their defects are most probably from NO_2 molecules, and they point out that the expected hyperfine structure from NO_3 would be much smaller than that observed. These authors also tried, unsuccessfully, to repeat Ard's results in sodium nitrate using γ- rather than X-irradiation. Both sets of authors also report other lines due to defects produced by the irradiation. In order to produce an NO_2 molecule by irradiation it is presumed that an O^- ion must be removed from an $(NO_3)^-$ group of the original crystal.

2.4.3. *Hydrogen atoms produced by irradiation.*

A survey of paramagnetic resonance from hydrogen atoms trapped in solids is given by Livingston, Zeldes and Taylor (1955). The atoms have been produced mainly by γ-irradiation at 77°K. Some typical results of these and other authors for the doublet hyperfine splitting and g-value can be summarized as follows :

Substance	$HClO_4$	H_2SO_4	H_3PO_4	H_2O†	Ice	U centre	Free $H_{(4)}$‡	Free H
Irradiation	γ	γ	γ	γ	γ	U.V.	None	None
$A/g\beta$ (Oe)	502	505	507	506	31	500	505	506
g	2·0022	2·0024	2·0024	2·00	2·00	2·00	2·00243	2·00230
Reference	(a)	(a)	(a)	(a)	(b)	(c)	(d)	(e), (f)

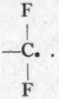
† H_2O absorbed on glass.
‡ Free H trapped at 4°K.

(a) Livingston *et al.* 1955 ; (b) Smaller, Matheson and Yasaitis 1954, Matheson and Smaller 1955 ; (c) Delbecq, Smaller and Yuster 1956 ; (d) Jen, Foner, Cochran and Bowers 1956 ; (e) Kusch 1955 ; (f) Beringer and Heald 1954.

Similar results have been obtained for all cases except $HClO_4$ and H_3PO_4 with D instead of H, where the splitting of the free atom triplet is $A/g\beta = 78$ Oe. The small splitting in ice (1/16 of the free atom value), is attributed by Smaller *et al.* to the effect of the solid on the electronic state of the H atom. Attempts by Livingston *et al.* to repeat these ice results at higher microwave frequencies (23 000 Mc/s as opposed to 350 Mc/s) have so far failed. On the other hand, Matheson and Smaller have used their 350 Mc/s spectrometer to measure the atomic H in irradiated H_2SO_4 and find results in good agreement with those of Livingston *et al.* All of these H atoms in solids have been measured at 77°K. The signal disappears as the sample is warmed up, due, it is thought, to combination of the H atoms to form H_2 molecules.

2.4.4. *Plastics.*

There have been a number of paramagnetic resonance measurements on irradiated plastics, of which we only mention two fairly typical examples, namely Fluon (otherwise known as Teflon) and Perspex. The structure of Fluon may be written $(CF_2)_n$. After x-irradiation Schneider (1955) finds a barely resolved triplet at $g = 2{\cdot}02$ and with $A/g\beta \simeq 19$ Oe. This is attributed to an unpaired electron located on the CF_2 end group of a broken polymer chain in which a C–C bond has been broken. The broken end then looks like



$$\overset{\displaystyle F}{\underset{\displaystyle F}{-C\cdot}}\ .$$

The hyperfine structure is thought to arise from the two ^{19}F nuclei, each with $I = 1/2$. On the other hand, Ard, Shields and Gordy (1955) find an eight line structure in irradiated fluon, with $g = 2{\cdot}0036$ and $A/g\beta \simeq 30$ Oe, which they attribute to the unpaired electron moving over seven equivalent ^{19}F nuclei. They also find an additional anisotropic triplet near $g = 2$, whose intensity grows at the expense of the eight line structure if the specimen is aged at room temperature for several days in an oxygen atmosphere. This they attribute to O_2 molecules which diffuse into the specimen and bond on to the radical at the end of the broken polymer chain. In x-irradiated Perspex, Schneider (1955) finds resonance absorption at $g = 2{\cdot}003$ with a nine-line hyperfine structure, $A/g\beta = 11$ Oe. He attributes this to two unpaired electrons with some exchange interaction between them. Each is supposed to be located on the CH_2 end group of a broken polymer chain, so it is suggested that four protons, each with $I = 1/2$, contribute to give the nine-line structure.

2.4.5. *Diamond.*

There are several examples of paramagnetic resonance in crystals which have suffered lattice damage as a result of heavy particle bombardment. We quote neutron irradiated diamond as a fairly typical case. The resonance spectrum is found (Griffiths, Owen and Ward 1955) to contain a strong central line at $g = 2{\cdot}0028$, and a number of weaker satellites which can be ascribed to magnetic centres with electronic spin $S = 1$ and with zero field splitting of the three spin levels into a doublet and singlet with separation $0{\cdot}01\ \text{cm}^{-1}$. Pryce has suggested that the central line might be due to single carbon vacancies and single interstitials, and the satellites to more complex aggregates of vacancies or interstitials, for example two interstitials might form a C_2 molecule (see O'Brien and Pryce 1955). The fact that the strong line is removed or greatly reduced by heating to 1000°c for a few minutes would then correspond to the interstitials diffusing back into vacant lattice sites. The satellite lines are not reduced by heat treatment and must correspond to some more stable defect such as a C_2 molecule. In general, it is to be expected that spectra due to lattice damage will be much more difficult to interpret exactly than, for example, those due to trapping centres.

For not only are several different types of defect likely to be present simultaneously, but also one cannot be sure that the original crystal structure is preserved in the neighbourhood of any particular defect.

2.5. *Transition Group Ion Impurities*

The paramagnetic resonance spectra from crystals containing ions with an unfilled d or f shell have been reviewed in a recent *Progress Report* (Bowers and Owen 1955), and no detailed account will be given here. A number of new results have been reported[†] which have been concerned mostly with the same ions as those used in earlier measurements, but present as impurities in different crystals. The spectra are then not fundamentally different from those found previously, but may show some interesting effect due to changes in crystal field or to changes in the bonding between the transition group ion and its neighbours. We quote only two examples. The first is MgO containing the substitutional impurity ions Cr^{3+}, V^{2+}, Mn^{2+}, Fe^{2+} and Ni^{2+}. One result shown by the paramagnetic resonance work (Low 1956) is that the crystal or molecular field acting on the ion, due to the nearest neighbour octahedron of oxygen ions, has the unusual property of accurate cubic symmetry. The second example is that of ZnF_2 containing the substitutional impurities Cr^{3+}, Mn^{2+}, Fe^{2+} and Co^{2+}. The interest in the spectra from these ions (Tinkham 1956) is that there is a ^{19}F hyperfine structure due to transfer of the unpaired d-electrons from the paramagnetic ion to orbits on the surrounding octahedron of F atoms, thus giving direct evidence for covalent bonding even in so ' ionic ' a complex. For Mn^{2+}, $3d^5$, Tinkham estimates that each of the unpaired electrons is in an orbit on a particular F atom for about 2% of the time. The hyperfine interaction with the ^{19}F nucleus also leads to a shift in the ^{19}F nuclear resonance frequency. Shulman and Jaccarino (1956) have measured this shift in MnF_2, and estimate an electron transfer in reasonable agreement with that found by Tinkham (see also Bleaney 1956).

§ 3. CONDUCTORS

In the following sections we shall outline some features of microwave absorption by conductors. We shall first consider electron spin resonance experiments on conduction electrons (e.g. in alkali metals), then spin resonance from paramagnetic impurities embedded in conductors (e.g. Mn in Cu), and finally non-resonant absorption of microwaves by metals at low temperatures (superconductors and anomalous skin effect).

3.1. *Spin Resonance from Conduction Electrons*

The main differences between spin resonance from conduction electrons, and that from paramagnetic ions in non-conducting materials are as follows : (i) The signal intensity is small, and approximately independent of temperature (apart from effects due to changes in skin depth), since only the electrons at the top of the Fermi distribution are unpaired. (ii) If the sample dimensions are larger than the skin depth ($\delta \sim 10^{-4}$ to 10^{-5} cm, for wavelengths ~ 1 cm), the amplitude and phase of the microwave radiation is different for magnetic carriers at different distances from the surface. The observed absorption line is then a mixture of the absorptive and dispersive parts of the susceptibility. (iii) The magnetic carriers are moving, and diffusing in and out of the skin depth. This further modifies the shape of the observed absorption line. Effects (ii) and (iii) can be eliminated by using a specimen in the form of a dispersion of particles small compared with the skin depth.

[†] A detailed table of these results is being prepared by Mr. J. Orton and will be published separately.

The first observation of conduction electron spin resonance was in sodium by Griswold, Kip and Kittel (1952). A general investigation has since been made of the resonance in Li, Na, K, Cs and Be (see, for example, Feher and Kip 1955, Levy 1956). These authors found no resonance in Mg, Al, Pd or W, possibly due to the limitations imposed by line width and the small signal intensity, though resonance in these and other metals had been reported by Al'tshuler, Kurenev and Salikov 1952, and Salikov 1952. The experiments have been made mostly with frequencies between 300 and 9000 Mc/s, and on specimens in the form of flat plates thick compared with the skin depth, or on fine particles small compared with the skin depth. The particle specimens have been either ultrasonically dispersed in paraffin wax, or produced by precipitation in frozen metal–ammonia solutions. For the case of flat plates, care must be taken to mount the specimen so that the E field in the cavity resonator is either normal to or zero at the surface of the plate, otherwise the losses are high and the Q of the cavity low. Apart from this precaution, the experimental techniques are similar to those for normal paramagnetic resonance.

An excellent short review of spin resonance from conduction electrons has been given by Kittel (1954). We now summarize the main experimental results.

3.1.1. *Signal intensity and susceptibility.*

One of the characteristic features of the resonance absorption is that the signal intensity is approximately independent of temperature, and in early work this was quoted as evidence that the observed signal was in fact due to conduction electrons. Measurement of this intensity gives the paramagnetic susceptibility of the conduction electrons more directly than any other available method. It has been accurately measured for Li and Na by Schumacher, Carver and Slichter (1954) and Schumacher and Slichter (1956). They compared the electron spin resonance line with the nuclear spin resonance from the Li or Na nuclei during the same experiment and using the same radio frequency. They find for Li $\chi_e = 2{\cdot}08 \times 10^{-6}$, and for Na $\chi_e = 0{\cdot}95 \times 10^{-6}$ c.g.s. volume units. These values show considerable divergence from the usual Pauli spin susceptibility theory, but are in reasonable agreement with the more refined theory of Pines and Bohm (1952) and Pines (1954). They are also in good agreement with values of χ_e estimated from the Knight shift found in nuclear resonance experiments on Li and Na.

3.1.2. *g-values.*

The observed g-values show that the conduction electrons behave to a first approximation like free electron-spins, though there is a small g-shift, $\Delta g = g_{\text{exp}} - 2{\cdot}0023$, which can be attributed to spin–orbit coupling effects. The experimental values are given in table 3.1. It can be seen that for the alkali metals they are in reasonable agreement with the theory of Argyres and Kahn (1954), Yafet (1952) and especially Brooks (1954). No detailed theory for Be appears to have been published.

Table 3.1. Summary of Electron Spin Resonance Results in Metals

	Li	Na	K	Cs	Be
g	2·0023(a)	2·0015(a) 2·0012 $\left\{\begin{array}{l}(300°\text{K})^{(c)}\\(77°\text{K})\end{array}\right.$ 2·0014 (4°K)	1·995(b)	1·93(b)	2·0032(a)
$\Delta g \times 10^4$ (expt)	$\leqslant \pm 1$(a)	-8 ± 2(a)	-70 ± 50(b)	-700 ± 200(b)	$+9 \pm 1$
(theory)	$-0·6$(d)	$-6·6$(e) $(-3·7)$(f)	-40(e)	-600(e)	
$b = d(\Delta H)/dT$ (Oe per 100°K)	—0·1(a)	2·5(a) 2(g)	11(b)		
Line width ΔH (Oe)	0·3 (77°K)(a) 0·3 (4°K) (temp. independent)	5·0 (200°K)(a, b) 0·1 (4°K) 3·6 (77°K)(g) 9·7 (385°K)	20 (140°K)(b) 5 (4°K)	13 (4°K)(b)	13(a) (temp. independent)

References: (a) Feher and Kip 1955, (b) Levy 1956, (c) Solt and Strandberg 1954, (d) Argyres and Kahn 1954, (e) Brooks 1954, (f) Yafet 1952, (g) Gutowsky and Frank 1954.

3.1.3. *Line widths and relaxation effects.*

It is found empirically that the line width (separation between maximum and minimum in the derivative of the absorption line) is either independent of temperature or is approximately of the form $\Delta H = a + bT$. Here a is a temperature independent term which increases with the impurity content of the sample, and bT is a contribution which increases linearly with the temperature T, and is approximately independent of the impurity content. Some of the experimental values are in table 3.1.

A discussion of some possible relaxation mechanisms which might contribute to the line width has been given by Overhauser (1953) (before any experimental results on spin resonance from conduction electrons were available). He concluded that the most important effect might be the interaction between the magnetic moment of one electron, and the field caused by the translational motion of a second electron. This leads to a temperature-dependent line width, but the numerical values turn out to be about two orders of magnitude smaller than the observed values.

A theory† which appears to be in better agreement with experiment is due to Elliott (1954), who considered the probability of an electron spin flip, via spin–orbit coupling, when there is a collision between a conduction electron and a thermal phonon or a lattice imperfection or impurity. This theory leads to a spin–lattice relaxation time

$$T_2 = \alpha T_R / (\Delta g)^2 \qquad \ldots\ldots(3.1)$$

where α is a numerical factor of order 1/30 if one considers only phonon collisions

---

† Further contributions to the theory have been made by Y. Yafet (unpublished, quoted by Feher and Kip 1955) and by D. Pines and C. P. Slichter (1955, *Phys. Rev.*, **100**, 1014).

(i.e. no imperfections), T_R is the usual collision relaxation time which can be found from resistivity measurements, and Δg is the g-shift which gives a measure of the spin–orbit interaction. For pure Na at 300°K, $T_R \simeq 3 \times 10^{-14}$ sec, and hence T_1 (theory) $\simeq 2 \times 10^{-9}$ sec. Feher and Kip find from their line width measurements T_1 (expt) $\simeq 9 \times 10^{-9}$ sec. The agreement is to be considered fairly reasonable in view of the approximations in the theory. The theory also predicts a linear dependence of ΔH on T for pure specimens as is observed. Finally, the observed temperature-dependent part of the line width is found to increase with Δg, i.e. with the spin–orbit interaction, as predicted, being too small to measure for Li, and increasing from Na to K. However, the present evidence does not show exact proportionality with $(\Delta g)^2$.

3.1.4. *Line shape.*

A typical feature of conduction electron spin resonance is the asymmetrical line shape found in specimens thick compared with the skin depth. The theory of this effect has been given by Dyson (1955), and verified in detail by the experiments of Feher and Kip (1955).

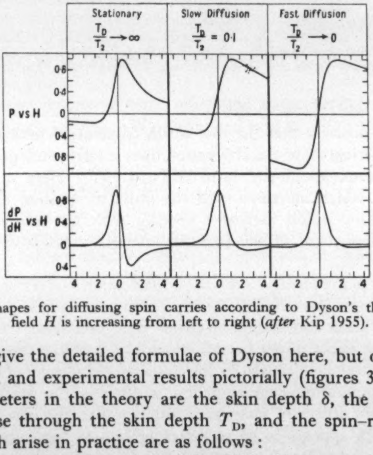


Figure 3.1. Line shapes for diffusing spin carries according to Dyson's theory. The magnetic field H is increasing from left to right (*after* Kip 1955).

We will not give the detailed formulae of Dyson here, but only indicate some of the theoretical and experimental results pictorially (figures 3.1 and 3.2). The important parameters in the theory are the skin depth δ, the time taken for an electron to diffuse through the skin depth T_D, and the spin–relaxation time T_2. Some cases which arise in practice are as follows :

(i) Dimensions of sample $\ll \delta$ (e.g. for Na at 300°K and 300 Mc/s, $\delta \simeq 6\mu$). For this case the amplitude and phase of the microwaves are the same throughout the sample. If the line width is entirely due to the relaxation time T_2 the power absorption P is of the form

$$P \propto \chi'' \propto \frac{1}{1+x^2} \qquad \ldots\ldots(3.2)$$

where $x = \gamma T_2(H-H_0)$, γ = gyromagnetic ratio, $H_0 = \omega/\gamma$ = resonance field, H = variable magnetic field. This is the usual symmetrical Lorentz line proportional to the absorptive part of the susceptibility χ'', such as might be found in normal paramagnetic resonance in a non-conductor.

(ii) Flat plates, with thickness $\gg \delta$, and with $T_D \gg T_2$. This is the case of stationary magnetic carriers suspended in the skin depth (e.g. nuclear resonance in metals, Bloembergen 1952, CuMn alloys, § 3.2), or of conduction electrons whose spins are flipped by a relaxation process before they have moved far in the skin depth and so they do not know they are diffusing. The power absorption is proportional to the real part of the surface impedance (cf. eqn (3.16)) $Z_{rf} \propto \mu^{1/2}(1+i)$. For the present case the complex permeability $\mu = 1 + 4\pi(\chi' - i\chi'')$, where $\chi' = x\chi''$ is the dispersive part of the susceptibility, and χ'' is given by equation (3.2). If $\chi', \chi'' \ll 1$, it follows that

$$P \propto \chi' + \chi'' + \text{constant}. \qquad \ldots\ldots(3.3)$$

This curve is plotted in figure 3.1.

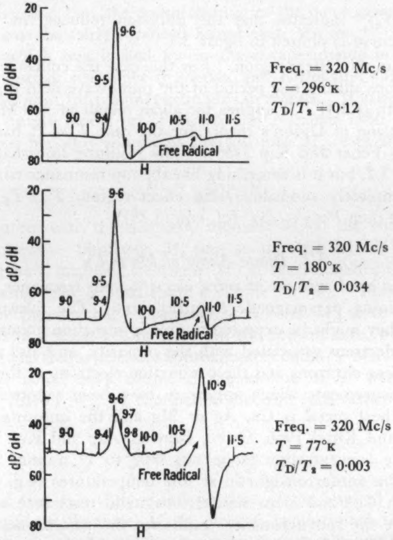


Figure 3.2. Electron spin resonance in thick plates of lithium for different ratios of T_D/T_2 as measured by Feher and Kip (1955). The markers are proportional to the magnetic field with 1 unit = 11·77 Oe (*after* Kip 1955).

(iii) Flat plates, with thickness $\gg \delta$ and $T_D \ll T_2$ (e.g. Li at 77°K where $T_D/T_2 \simeq 0{\cdot}003$ and $T_2 \simeq 3{\cdot}5 \times 10^{-8}$ sec, see figure 3.2). The electrons diffuse in and out of the skin depth many times before their spin is flipped. To estimate the complex permeability in the wave impedance one has in effect to estimate the average magnetization in the skin depth of the diffusing electrons. Very roughly, the result is that the observed power absorption has the form of a dispersion curve $P \sim \chi'$. More precisely, the absorption line has wings of low intensity spread out over a width of the order of $1/T_D$, and the observed central part of the line is of the form

$$P \propto \left(\frac{T_D}{T_2}\right)^{1/2} \left[\frac{(1+x^2)^{1/2}-1}{1+x^2}\right]^{1/2} \frac{x}{|x|}. \qquad \ldots\ldots(3.4)$$

The factor $(T_D/T_2)^{1/2}$ indicates that the diffusion reduces the observed signal intensity. This curve is plotted in figure 3.1.

(iv) Anomalous skin effect region. For this case the collision free path of the conduction electrons during one period of the microwave field is greater than the classical skin depth (§ 3.4). It applies for alkali metals at 300 Mc/s below about 40°K. The extension of Dyson's theory for the case $T_D \ll T_2$ has been made by Kittel (quoted by Feher and Kip 1955). The resulting line shape is not shown in figures 3.1 and 3.2, but it is something like the top resonance curve of figure 3.2.

If, in the completely anomalous skin effect region, $T_D \gg T_2$ (e.g. stationary magnetic carriers) then $P \propto \chi' + \sqrt{3}\chi''$ (cf. eqn (3.25)).

3.2. *Dilute Alloys of Mn in Cu*

We now give a brief account of some electron spin resonance experiments on conductors containing paramagnetic impurity ions. The measurements are of interest because they might be expected to give information about the state of the bound unpaired electrons associated with the impurity, and also about the interaction between these electrons and the conduction electrons of the host metal.

The only measurements which appear to have been reported so far are for alloys where the host metal is Cu, Ag or Mg and the impurity is Mn (Owen, Browne, Knight and Kittel 1956, Owen, Browne, Arp and Kip 1957). For the CuMn system the concentration range was 0·03 to 11 atomic % Mn. These alloys appear to be antiferromagnetic at low temperatures (e.g. below 37°K for 5·6 at. % Mn in Cu) and show antiferromagnetic resonance absorption (§ 9). Here we only give the spin resonance results for the paramagnetic region which are as follows : (i) There is a single absorption line with close to free spin *g*-value, $g = 2{\cdot}00 \pm 0{\cdot}01$. (ii) Single crystal measurements show that there is negligible anisotropy. (iii) The line shape is consistent with that expected from fixed magnetic carriers suspended in the skin depth (eqn (3.3)). (iv) The line width has minimum value of about 100 oersteds, and increases linearly with temperature, corresponding at 300°K to a relaxation time $T_2 \simeq 10^{-10}$ sec. These results appear to be consistent with the Mn being in an S state, such as might result from Mn^{2+}, $3\mathrm{d}^5$, ${}^6\mathrm{S}_{5/2}$ with spin $S = 5/2$. However, susceptibility measurements indicate that there may be rather less than five unpaired spins per Mn atom.

The detailed magnetic properties of these alloys are not yet fully understood. A mechanism which might be expected to affect the properties is an s–d exchange coupling

$$\mathscr{H} = -A\mathbf{S}.\mathbf{s} \qquad \ldots\ldots(3.5)$$

where $S = 5/2$ is the spin of the $3\mathrm{d}^5$ Mn ion core, and $s = 1/2$ is the spin of a 4s conduction electron. This could lead to two effects in the spin resonance spectrum from the Mn^{2+} ions, (i) a short relaxation time, (ii) a *g*-shift. Experimentally, the measured relaxation time is in fact of order 1000 times shorter than that found for Mn^{2+} ions in ionic crystals, and leads to a value $A_{\exp} \lesssim \frac{1}{8} A_0$, where $A_0 = 3160\ \mathrm{cm}^{-1}$ is the coupling constant estimated from free ion spectroscopic data. The absence of an observed *g*-shift leads to $A_{\exp} < \frac{1}{8} A_0$. Other experiments (see references above) including antiferromagnetic resonance, nuclear resonance, and susceptibility, also suggest that the s–d coupling is small, perhaps 1/10 to 1/20 of the free ion value, and that the magnetization of the conduction electrons by the s–d interaction may be fairly localized round each Mn ion.†

The *g*-shift which was looked for in these experiments is analogous to the Knight shift in nuclear resonance in metals, and Kittel first suggested that there might be a similar shift for paramagnetic ions in metals. It arises in the following way. Using a simple model (cf. Fröhlich and Nabarro 1940, Zener 1951), the magnetic moment per unit volume of the Mn ions M_A, sees an internal field

$$H' = \frac{A}{4\beta^2 N_e} M_e \qquad \ldots\ldots(3.6)$$

due to the coupling with the magnetic moment M_e of the conduction electrons (N_e per unit volume). Similarly, M_e sees an internal field $H'' = (A/4\beta^2 N_e) M_A$. There will then be a magnetic field shift H' in the resonance from M_A provided that the precessing field H' can exert a couple on M_A. This could happen if, for example, the precessional frequencies of H' and M_A were very different (cf. the Knight shift in nuclear resonance). In the present case, where it is expected that the precessional frequencies are approximately equal (*g*–Mn ions $\simeq$ *g*-conduction electrons), there could be a couple if the motion of H' were heavily damped compared with that of M_A, by conduction electron spin–lattice interaction. The fractional *g*-shift is then

$$\frac{\Delta g}{g} = \frac{H'}{H} = \frac{A}{4\beta^2 N_e} \frac{M_e}{H}. \qquad \ldots\ldots(3.7)$$

It will be noted that this shift is independent of the applied field H and of the microwave frequency, because M_e is directly proportional to H. If equation (3.7) is applied to ferromagnetic resonance on the other hand (Kittel and Mitchell 1956, see § 7), M_e is practically constant at a given temperature since it is mainly determined by the large constant internal field H'' arising from the now saturated magnetization M_A. The *g*-shift is then inversely proportional to the applied field and to the frequency used in the experiment.

3.3. *Superconducting Metals*

In the following sections we shall outline some features of electrical conduction at microwave frequencies which are of interest in connection with the properties of metals at low temperatures. Detailed reviews of this field have been given by Sondheimer (1952) and Pippard (1954) and we shall limit our report to an elementary account of some of the more interesting results.

---

† *Note added in proof*: Recent theoretical work by E. W. Hart (1957, *Phys. Rev.*, **106**, 467) and K. Yosida (1957, *Phys. Rev.*, in the press) has shown that the conduction electron magnetization is likely to be localized. A comparison of the experimental results with Yosida's theory gives an s–d interaction of the order A_0, the free ion value.

Superconducting metals show interesting properties in the microwave region of the spectrum. The transition to the superconducting state is a second-order process without any associated latent heat and only appears as a true discontinuity in the resistance for the limiting case of steady fields and currents. At infra-red frequencies the metal possesses a finite resistance equal to that which it would have as a normal conductor (London 1940) and there is a region at the short wavelength end of the microwave spectrum where the metal loses some but not all of its resistivity in the superconducting state. The resistivity at these frequencies may be derived from measurements of the Q value for a cavity resonator at various temperatures (see, for example, Pippard 1954) or from the transmission properties of metallic films (Glover and Tinkham 1956). The recent measurements of Blevins *et al.* (1955) and Biondi *et al.* (1956) using cavity resonator techniques suggest that both the temperature for the onset of the superconducting phase and the resistivity are frequency dependent in the case of tin and aluminium whilst Glover and Tinkham (1956) found an anomalously high transmission through lead films in the long wavelength infra-red.

A qualitative explanation of the change in resistivity can be given by the ' two fluid ' model of a superconductor which supposes that in such a metal the conduction electrons can occupy either of two states, in one of which they have no resistance, whereas in the other they behave as normal conducting electrons. The fraction of superconducting electrons increases steadily from zero to unity as the temperature falls below the transition towards the absolute zero. For the d.c. case the observed resistance is zero at all temperatures below the transition because the super-electrons will short circuit the normal electrons, but in the a.c. case the super-electrons will present a reactive impedance because of their inertia and some of the current will be carried by the normal electrons giving rise to a power dissipation which will be dependent on frequency and temperature (since it will depend upon the number of normal electrons).

For alternating electric fields the situation can be described macroscopically by writing the total current density $\mathbf{J}$ as the sum of two terms, $\mathbf{J}_n$ arising from the normal electrons and $\mathbf{J}_s$ arising from the superconducting electrons :

$$\mathbf{J} = \mathbf{J}_n + \mathbf{J}_s. \qquad \ldots\ldots(3.8)$$

The assumption of zero d.c. resistivity for the electrons contributing to J_s suggests an equation of motion with no damping term

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E} \qquad \ldots\ldots(3.9)$$

and, since $\mathbf{J}_s = -n_s e\mathbf{v}$, with n_s equal to the number of electrons per unit volume in the superconducting state,

$$\frac{d\mathbf{J}_s}{dt} = \frac{n_s e^2}{m}\mathbf{E} = \mathbf{E}/\lambda. \qquad \ldots\ldots(3.10)$$

The equation which replaces Ohm's law for a superconducting metal is then

$$\dot{\mathbf{J}} = \sigma_n\dot{\mathbf{E}} + \frac{1}{\lambda}\mathbf{E} \qquad \ldots\ldots(3.11)$$

where σ_n is the conductivity arising from the 'normal' electrons which give rise to Joule heating. This is not the same as the normal conductivity measured at the same temperature when the superconductivity is destroyed by a magnetic field, since in this case the number of normal electrons is increased by a contribution coming from the former superconducting electrons. On an energy gap model of a superconductor σ_n will be both temperature and frequency dependent. If the time variation of the high-frequency field is of the form $\exp(j\omega t)$, we obtain from equation (3.11)

$$\mathbf{J} = \left(\sigma_n + \frac{1}{j\omega\lambda}\right)\mathbf{E} \equiv \sigma_s \mathbf{E}. \qquad \ldots\ldots(3.12)$$

On this model Maxwell's equations retain the same form as in the case of a normal metal and the usual relations describing the electromagnetic properties of conductors may be taken over by writing an effective conductivity $\sigma_s = \sigma_n + 1/j\omega\lambda$. It has been customary to neglect the displacement term in Maxwell's equations, although recently arguments in favour of a large effective dielectric constant have been put forward by Guisburg and Landau (1950) and by Tinkham (1956) so that there may be a significant contribution. For our purpose, however, we may assume that the displacement term can be neglected. The power loss in the metal and its reflection coefficient can then be derived from the wave impedance in the form

$$Z = (j\omega\mu/4\pi\sigma)^{1/2} \qquad \ldots\ldots(3.13)$$

giving

$$Z_s = \left(\frac{j\omega\mu}{4\pi\sigma_s}\right)^{1/2} = \left\{j\omega\mu/4\pi\left(\sigma_n + \frac{1}{j\omega\lambda}\right)\right\}^{1/2}, \qquad \ldots\ldots(3.14)$$

that is

$$Z_s = \left(\frac{\omega\mu}{8\pi|\sigma_s|^2}\right)^{1/2}\left\{\left(|\sigma_s| - \frac{1}{\omega\lambda}\right)^{1/2} + j\left(|\sigma_s| + \frac{1}{\omega\lambda}\right)^{1/2}\right\}. \qquad \ldots\ldots(3.15)$$

At sufficiently high frequencies $\sigma_s \to \sigma_n$, $1/\omega\lambda \to 0$ and we have

$$Z_{\mathrm{hf}} = \left(\frac{\omega\mu}{8\pi\sigma_n}\right)^{1/2}(1+j) \qquad \ldots\ldots(3.16)$$

which is the familiar classical description of a normal metal whilst at low frequencies

$$Z_{\mathrm{lf}} = j\left(\frac{\omega^2\mu\lambda}{4\pi}\right)^{1/2} \qquad \ldots\ldots(3.17)$$

which is characteristic of a superconducting material. Evidently the transition from low-frequency to high-frequency behaviour occurs at a frequency of the order

$$\omega_t \sim \frac{1}{\sigma_n\lambda} = \frac{n_s}{n_n}\frac{1}{\tau} \qquad \ldots\ldots(3.18)$$

where τ is the collision time on a simple free electron model. A more revealing expression for the critical frequency ω_t is obtained from the relation between wave impedance and the penetration depth or skin depth δ:

$$Z = \frac{c}{4\pi}(\sigma\Delta)^{-1} \qquad \ldots\ldots(3.19)$$

and $1/\delta = \mathscr{R}(1/\Delta)$. We obtain

$$\delta_{\mathrm{hf}} = (c^2/2\pi\omega\mu\sigma_n)^{1/2}, \quad \delta_{\mathrm{lf}} = (c^2\lambda/4\pi\mu)^{1/2}. \qquad \ldots\ldots(3.20)$$

Thus the critical frequency corresponds to a high-frequency skin depth of the same order of magnitude as the low-frequency penetration depth. This latter is of the order 10^{-5} cm and so the transition occurs in the microwave range.

The temperature variation of resistance for superconducting tin is shown in figure 3.3 for various frequencies. The macroscopic theory outlined above gives a rough account of the measurements in so far as the infra-red resistance corresponds to the normal impedance and there is some residual resistance at microwave frequencies below the superconducting transition. There are, however, additional effects in tin (Blevins *et al.* 1955) where the transition temperature itself is apparently frequency dependent, and in lead where there is an anomalous transparency in the long wavelength infra-red (Glover and Tinkham 1956). A frequency dependent transition temperature has also been observed in aluminium (Biondi *et al.* 1956) together with some additional unexplained changes in the superconducting state where the residual resistance does not vary monotonically with temperature.

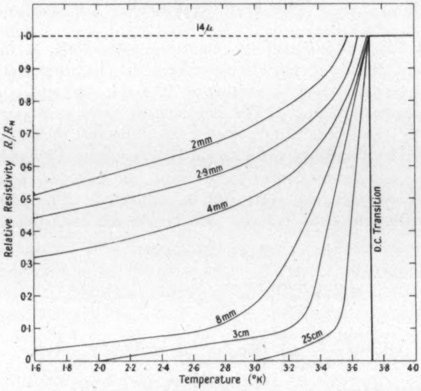


Figure 3.3. Temperature variation of the resistivity for superconducting tin, based on Pippard (1954) and Blevins *et al.* (1955).

Buckingham (1956) has suggested that an explanation of the change in transition temperature with frequency may be sought on a quantum model of a superconductor for which there is an energy gap ϵ_0 between the normal and superconducting states at the absolute zero, but this energy gap decreases with increasing temperature and eventually vanishes at the d.c. transition temperature T_c. If the properties of the normal metal are retained for frequencies such that $\hbar\omega > \epsilon_T$, then the superconducting transition will not occur until the temperature has fallen to T_ω, with

$$T_c - T_\omega \propto \omega^2 \qquad \ldots\ldots(3.21)$$

near the d.c. transition temperature. Such a relation seems to fit the present data for tin but obviously more experimental and theoretical work is needed.

Tinkham (1956) has suggested that the anomalous transparency of superconducting lead films in the long wavelength infra-red may arise from the frequency dependence of the conductivity σ_s. If σ_s is written in the form $\sigma_s = \sigma_1(\omega) - j\sigma_2(\omega)$ the ratio of power transmitted through the superconducting film T_s to that transmitted through the normal film above the transition point T_N is of the form

$$\frac{T_s}{T_N} = \left[\left\{T_N{}^{1/2} + (1 - T_N{}^{1/2})\frac{\sigma_1}{\sigma_N}\right\}^2 + \left\{(1 - T_N{}^{1/2})\frac{\sigma_2}{\sigma_N}\right\}^2\right]^{-1} \qquad \ldots\ldots(3.22)$$

where σ_N is the conductivity of the normal metal corresponding to T_N. At low frequencies σ_2 is predominant and σ_1 is small so that the material is a typical superconductor with T_s much less than T_N. At high frequencies, on the other hand, $\sigma_2 \to 0$ and σ_1 is predominant but here $\sigma_1 \to \sigma_N$ and so $T_s \to T_N$ and the material behaves like a normal metal. At an intermediate frequency there may be a situation in which both σ_1 and σ_2 are small. In this case T_s/T_N may be of the order $T_N{}^{-1/2}$ and since T_N is always much less than unity $T_s > T_N$ and the material is more transparent in the superconducting state than as a normal metal.

3.4. *Anomalous Skin Effect*

In the discussion so far, the 'normal' conduction electrons have been regarded as possessing the usual properties ascribed to current carriers which obey Ohm's law. It is customary to describe the relaxation process by which such electrons exchange energy with the crystal lattice in terms of a characteristic time τ which represents the fact that any additional momentum acquired by an electron from the field will decay at a rate $\exp(-t/\tau)$. Ohm's law is only valid when the electric field which acts on the electrons is a spatial and temporal constant throughout the motion of the electron between collisions. Effects arising from a temporal variation of the electric field (relaxation effects) do not become significant until $\nu \sim \tau^{-1}$ which corresponds to the long wavelength infra-red but anomalous skin effects due to the spatial variation of the electric field are readily observed at microwave frequencies.

When the distance l_0 traversed by an electron between collisions becomes greater than the skin depth, the current density at a point $\mathbf{r}_0$ will be determined by contributions from electrons which will have been subject to widely differing electric fields. In travelling towards the point $\mathbf{r}_0$, however, the momentum acquired from the electric field at $\mathbf{r}$ will have decayed due to collisions by a factor $\exp(-t/\tau) = \exp\{-|\mathbf{r} - \mathbf{r}_0|/l_0\}$ so that the additional momentum in excess of the Fermi momentum, which will be contributed by electrons arriving at $\mathbf{r}_0$ will be of the form

$$\mathbf{M} = \frac{e}{v_0}\int_{r_0}^{\infty} E(r)\exp\{-|r - r_0|/l_0\}\,d|r - r_0| \qquad \ldots\ldots(3.23)$$

in which v_0 is the velocity at the Fermi surface.

The current density at the point $\mathbf{r}_0$ is then given by the relation (Pippard 1954)

$$\mathbf{J}(\mathbf{r}_0) = \frac{3ne^2}{mv_0}\int_{r_0}^{\infty}\frac{\mathbf{R}[\mathbf{R}\,.\,\mathbf{E}(\mathbf{r})]\exp(-|\mathbf{R}|/l_0)}{|\mathbf{R}|^2}\,d\mathbf{R} \qquad \ldots\ldots(3.24)$$

$$\mathbf{R} = \mathbf{r} - \mathbf{r}_0$$

which represents the contributions to the current density at $\mathbf{r}_0$ from electrons which have arrived from different distances and in different directions. Equation (3.24) reduces to the familiar relation

$$\mathbf{J} = (ne^2 l_0/mv_0)\,\mathbf{E} = \sigma\mathbf{E}$$

when the electric field is constant throughout the region with dimensions l_0 and is evidently a generalization of Ohm's law.

As a consequence of the spatial variation of the electric field, the surface impedance Z differs from that which is characteristic of a normal Ohm's law conductor. In the limiting case when the collision parameter l_0 is very large compared with the skin depth for the high frequency field, Z_∞ is given by (Reuter and Sondheimer 1948)

$$Z_\infty = \alpha(1+j\sqrt{3})\,(\omega^2 l_0/\sigma)^{1/3}. \qquad \ldots\ldots(3.25)$$

In this equation α is a numerical constant whose value is determined by the nature of the scattering of the conductions at the metal surface. For specular reflection at the surface $\alpha = 1{\cdot}563$, whereas for diffuse scattering $\alpha = 1{\cdot}759$. Since Z_∞ depends upon the ratio l_0/σ, the surface impedance is independent of the d.c. conductivity σ; this behaviour is in contrast to the $\sigma^{-1/2}$ dependence of an Ohm's law conductor for which $Z = (\omega\mu/8\pi\sigma)^{1/2}(1+j)$. Furthermore, the resistive and reactive components are no longer equal but should be in the ratio $1:\sqrt{3}$. Chambers' (1952) measurements on copper, silver, gold, lead, tin, cadmium and aluminium indicate that the conduction electrons undergo diffuse scattering at a metal surface, whilst Pippard (1947, 1950) has confirmed that Z_∞ varies as $\omega^{2/3}$ from measurements on tin at 25 cm and 3 cm wavelength. From measurements on superconducting tin he also derived the ratio X_∞/R_∞ to be 1·76 in good agreement with the predicted value of $\sqrt{3}$ where $Z_\infty = R_\infty + jX_\infty$.

Determinations of Z_∞ may be used to derive the value of l_0/σ from which, in principle, the number of conduction electrons per unit volume may be obtained. Thus for a simple free electron model $l_0/\sigma = (mv_0/Ne^2)$. An experimental determination of the resistive component of Z_∞ may be obtained from the measured Q of a resonant cavity and measurements by Chambers (1952) have shown that the simple free electron model gives values for the number of conduction electrons per atom near to the expected value of unity in the case of the metals Cu (1·01), Ag (0·61) and Au (0·60). However, these are probably rather rough estimates since the final results depend upon $\mathscr{R}Z_\infty^{3/2}$ and so are very sensitive to any surface strains which will modify the impedance. Chambers (1952) has pointed out that a more significant factor which can be derived from the ratio l_0/σ is the area A of the Fermi surface in momentum space. He derives $\sigma/l_0 = A(2e^2/3h^3)$ for measurements on polycrystalline specimens which give an average over all crystal orientations, and it is suggested that this relation is true for the complicated Fermi surfaces applicable to real metals as well as for the simple free electron gas, provided that A does not include those regions of the Fermi surface which coincide with Brillouin zone boundaries. On the other hand, measurements on single crystals may be of importance in the investigation of the shape and orientation of the Fermi surface. Investigations by Pippard (1950) on tin at 3 cm indicate a more complex anisotropy than is found in Ohm's law conductors, and the resistive component of Z_∞ depends not only upon the direction of current flow but also upon the orientation of the metal surface relative to the crystal axis.

§ 4. ELECTRICAL ABSORPTION IN SEMICONDUCTORS

4.1. *General Discussion*

In this section we shall discuss the absorption of microwave radiation in semiconductors which arises from the translational motion of the charge carriers. The most important physical aspect is the ' cyclotron ' resonant absorption which provides a direct determination of the constants in the equations for the energy surfaces describing the motion of charge carriers. Here the microwave electric field induces transitions between the allowed (orbital) states for a carrier moving in a magnetic field. In classical language this is described as an anomaly in the high-frequency electrical conductivity, a description which is analogous to the discussion of electron spin resonance in terms of the high-frequency magnetic susceptibility. Quantum mechanical discussions have been given for ideal and for more practical cases (Dorfman 1951, Dingle 1952, Luttinger and Kohn 1954, 1955). In this report, however, we shall use semi-classical methods (Shockley 1953, Dresselhaus, Kip and Kittel 1955 a, Lax, Zeiger and Dexter 1954) to describe the phenomena. The general conclusions are the same in both the quantum and classical derivations, but there are some additional absorption lines arising from degenerate bands in the quantum mechanical theory. These have been observed by Fletcher *et al.* (1955).

A free charge q moving in a magnetic field $\mathbf{H}$ will spiral around the field direction with a natural frequency which is determined by the equation of motion $m(d\mathbf{v}/dt) = (q/c)(\mathbf{v} \times \mathbf{H})$. Taking the z axis along the direction of $\mathbf{H}$ this equation may be written in component form as

$$m\dot{v}_x = (q/c)\,v_y H, \quad m\dot{v}_y = -(q/c)\,v_x H, \quad m\dot{v}_z = 0$$

and these equations can be solved for v_x, v_y. For example,

$$m\ddot{v}_x = (q/c)\,\dot{v}_y H = -(q^2H^2/mc^2)\,v_x \quad \text{so that} \quad v_x = A\sin(\omega t+\phi)$$

with $\omega = qH/mc$; similarly, $v_y = A\cos(\omega t+\phi)$. The charge, therefore, moves in a path having a circular projection on a plane perpendicular to the magnetic field and the angular frequency of rotation is $\omega = qH/mc$. If an electric field of this frequency is applied to the charge in the plane perpendicular to the magnetic field, the charge will acquire energy and this will result in a ' cyclotron ' resonant absorption provided that the charge can perform a reasonable number of orbits before being thrown out of phase with the electric field by a scattering collision. This type of absorption process is an electric dipole transition in distinction from a magnetic dipole transition which occurs in electron spin resonant absorption. The sensitivity of a microwave spectroscope for this cyclotron resonance may be compared with the sensitivity for spin resonance under equivalent conditions of, for example, line width, effective number of carriers, and measuring frequency. The energy density in a microwave cavity is the same at a position where the electric field is a maximum as at a position where the magnetic field is a maximum, and so the relative sensitivity is the ratio of the transition probabilities, or the squares of the effective dipole moments $(e\bar{r}/\beta)^2$. For a typical case (Dresselhaus *et al.* 1955 a) the orbital radius of an electron with thermal velocity at 4°K, in a field of 1000 oersteds and effective mass one-tenth of that of a free electron, is about 5×10^{-5} cm, so that $(e\bar{r}/\beta)^2\sim10^{12}$. The cyclotron resonance absorption is therefore extremely intense compared with the electron spin absorption, but the full gain in sensitivity is not usually achieved because the line widths are of the order of hundreds of gauss. Under normal conditions about 10^8 carriers per cm^3 can be detected and so it is possible to investigate the properties of the band edge in a semiconductor whilst the material is effectively an insulator.

The translational motion of the electrons and holes at the band edge in an actual material is governed by the available energy states. The density of these states will not in general correspond to the free electron model and it has been customary to characterize the motion of a charge carrier by an 'effective mass' parameter which is related to the constants in the expression for the energy when written as a function of the characteristic wave vector **k**. This is essentially a generalization of the Newtonian force–momentum relation and the effective mass m^* is given in terms of the energy $\epsilon(\mathbf{k})$ by the relation (Seitz 1940)

$$\frac{1}{m^*}=\frac{1}{\hbar^2}\operatorname{grad}_{\mathbf{k}}\operatorname{grad}_{\mathbf{k}}\epsilon(\mathbf{k}), \qquad \ldots\ldots(4.1)$$

whilst the velocity of the charge carrier is derived from the energy by the relation (Seitz 1940)

$$\mathbf{v}=\frac{1}{\hbar}\operatorname{grad}_{\mathbf{k}}\epsilon(\mathbf{k}). \qquad \ldots\ldots(4.2)$$

The force acting on a charge carrier is related to its wave vector by the relation $\mathbf{F}=\hbar\dot{\mathbf{k}}$ and its motion in an electric and magnetic field is therefore described by the equations

$$\hbar\dot{\mathbf{k}}=\mathbf{F}=-e\left(\mathbf{E}+\frac{1}{c}\mathbf{v}\times\mathbf{H}\right)-\nu\hbar\mathbf{k} \qquad \mathbf{v}=\frac{1}{\hbar}\operatorname{grad}_{\mathbf{k}}\epsilon(\mathbf{k}). \qquad \ldots\ldots(4.3)$$

Here the term $\nu\hbar\mathbf{k}$ is introduced to take account of collisions by analogy with a classical damping term. A true resonance phenomenon at a measuring frequency ω will only be observed if this term is sufficiently small, i.e. $\omega/\nu\gg1$. This implies that the collision time is long compared with the microwave period. Equations (4.3) reduce straightforwardly to the relations for a free charge which have been discussed above, when damping is neglected. In this case the relations giving the energy ϵ in terms of the wave vector **k** are $\epsilon=\frac{1}{2}mv^2$, $\mathbf{k}=2\pi/\lambda=m\mathbf{v}/\hbar$. In a solid the relation between the wave vector **k** and the energy ϵ depends upon the symmetry of the crystal lattice structure and upon the potential field due to the positive ion cores. The energy relations are too complicated to derive rigorously and approximations which reduce to a parabolic form in limiting cases are usually given. The surfaces of constant energy when plotted as functions of **k** in three dimensions are frequently ellipsoids and these will be so oriented as to preserve the symmetry of the crystal lattice. The resonance spectrum which will be observed for an arbitrary orientation of the magnetic field may then exhibit a number of absorption lines arising from the differently oriented energy surfaces for both electrons and holes. An example of such a spectrum from germanium is shown in figure 4.1 (Dresselhaus *et al.* 1955 a).

4.2. *Experimental Techniques*

The cyclotron resonance arises from the interaction of the charge carriers with the component of the microwave electric field which is perpendicular to the steady magnetic field.† Except for this requirement the experimental techniques are similar to those commonly used in electron spin resonance and the same apparatus may conveniently be employed for both types of experiment. Thus if an H_1 mode resonant cavity is used, both spin and cyclotron resonance may be observed if it is possible to rotate the steady magnetic field through 90° about the axis of the cavity. Moreover, owing to the large dielectric constant of germanium (~16) and silicon (~12) and because of the high conduction losses which often occur at temperatures above about 20°K, it is usually quite satisfactory as well as advantageous to mount the sample on the end wall of the cavity in the same way as for spin resonance. The measurements reported in the literature (for example, Lax *et al.* 1954, Dresselhaus *et al.* 1955 a, Fletcher *et al.* 1955, Dexter *et al.* 1956) have been carried out at liquid helium temperatures where most of the carriers are trapped. It is, however, possible to observe cyclotron resonance in germanium at 20°K and, when the germanium is doped with suitable impurities, for example gold, the resonance may be observed at temperatures above 90°K (Bagguley, Powell and Taylor 1957). Plasma resonance in indium antimonide has been observed at 4°K and 77°K (Dresselhaus *et al.* 1955 b).

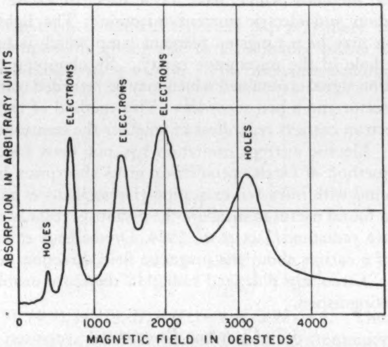


Figure 4.1. Typical cyclotron resonance results in germanium near 24 000 Mc/s and 4°K: direct copy from a recorder trace of power absorption against static magnetic field in an orientation in a (110) plane at 60° from a [100] axis (*after* Dresselhaus *et al.* 1955 a).

---

† Resonance effects have also been detected from the component of microwave electric field parallel to the applied magnetic field (Lax *et al.* 1954).

In spin resonance experiments an audio-frequency absorption signal is usually obtained by magnetic field modulation. This method may also be applied directly to the cyclotron resonance experiments or the reflection klystron itself may be modulated. The microwave field or thermal energy may be used to free the carriers from the bound states. This type of excitation usually produces carriers of one type only, electrons if the semiconductor is n-type and holes when the semiconductor is p-type. Microwave ionization is very convenient for locating the position of the absorption peaks but the measured line widths are very dependent upon the power level which is used. This arises from the carrier multiplication which may occur within an absorption line where the charge carriers can pick up more energy from the microwave field. The peak height of the absorption line is considerably increased and there is an apparent reduction in line width. Other methods which have been used to excite carriers include infra-red illumination and electric current injection. The light source for the infra-red excitation may be a tungsten filament lamp which is focused on to the sample through a hole in the microwave cavity. By chopping the light beam a modulated absorption signal is obtained which may be recorded using a narrow band phase sensitive detector and a pen recorder. This method of excitation produces both holes and electron carriers regardless of whether the semiconductor is initially n-type or p-type. Electric current excitation has not been used so widely. It appears that this method of carrier production gives absorption lines with greater widths than are found with infra-red excitation (Dresselhaus *et al.* 1955 a).

It has also been found useful to measure the cyclotron resonance with circularly polarized microwave radiation (Lax *et al.* 1954, Dresselhaus *et al.* 1955 a). The sense of rotation of a carrier about the magnetic field direction depends upon the sign of its charge. Circularly polarized radiation therefore enables the electrons and holes to be distinguished.

4.3. *Cyclotron Resonance*

The semiconductors germanium and silicon have been examined in greatest detail. The energy surfaces at the bottom of the conduction band are believed to be of the form

$$\epsilon(\mathbf{k}) = \frac{\hbar^2}{2}\left(\frac{k_x^2 + k_y^2}{m_1} + \frac{k_z^2}{m_2}\right). \qquad \ldots\ldots(4.4)$$

At the top of the valence band, the corresponding equation is

$$\epsilon(\mathbf{k}) = Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{1/2}. \qquad \ldots\ldots(4.5)$$

The constants m_1, m_2 of equation (4.4) are directly related to the effective mass parameters m^* for the x, y and z directions by equation (4.1). The terms B and C in equation (4.5) are perturbation terms arising from coupling between different energy bands. When the energy gap between the bands is sufficiently large B and C tend to zero and A reduces to $\hbar^2/2m$. In this case equation (4.5) describes a carrier with an isotropic mass equal to that for a free electron.

For the electrons at the bottom of the conduction band equations (4.3), (4.4) give an equation of motion

$$\frac{d}{dt}(\mathbf{m}.\mathbf{v}) + \nu\mathbf{m}.\mathbf{v} = -e\left[\mathbf{E} + \frac{1}{c}(\mathbf{v} \times \mathbf{H})\right] \qquad \ldots\ldots(4.6)$$

in which $\mathbf{m}$ is the effective mass tensor

$$\mathbf{m} \equiv \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_1 & 0 \\ 0 & 0 & m_2 \end{pmatrix}. \qquad \ldots\ldots(4.7)$$

In component form, with $\mathbf{E} = \mathbf{E}\exp(j\omega t)$ and $\mathbf{H}$ in the xz plane, we obtain

$$\begin{gathered} (\nu + j\omega m_1)\, v_x + \frac{eH_z}{c} v_y = -eE_x \\ -\frac{eH_z v_x}{c} + (\nu + j\omega m_1)\, v_y + \frac{eH_x}{c} v_z = -eE_y \\ -\frac{eH_x}{c} v_y + (\nu + j\omega m_2)\, v_z = -eE_z. \end{gathered} \qquad \ldots\ldots(4.8)$$

When the damping term is sufficiently small ($\omega/\nu \gg 1$) there will be an anomaly in the velocity components giving rise to a resonance power absorption because of the change in conductivity $\sigma = Nev$. The resonance condition is

$$\begin{vmatrix} j\omega m_1 & \dfrac{eH_z}{c} & 0 \\ -\dfrac{eH_z}{c} & j\omega m_1 & \dfrac{eH_x}{c} \\ 0 & -\dfrac{eH_x}{c} & j\omega m_2 \end{vmatrix} = 0 \qquad \ldots\ldots(4.9)$$

that is

$$\frac{\omega^2 c^2}{e^2 H^2} = \left(\frac{\cos^2\theta}{m_1{}^2} + \frac{\sin^2\theta}{m_2 m_1}\right) = \left(\frac{1}{m^*}\right)^2. \qquad \ldots\ldots(4.10)$$

In equations (4.8), (4.9), (4.10), $H_z \equiv H\cos\theta, H_x \equiv H\sin\theta$. Thus from the angular variation of the resonance field for a fixed measuring frequency the orientation of the principle axes of the energy surface may be obtained together with the constants m_1 and m_2 (see table 4.1). A typical set of experimental points together with the theoretical curve from equation (4.10) is shown in figure 4.2 (Dresselhaus, Kip and Kittel 1955 a). Here the effective mass parameter defined by the relation $m^* = eH/\omega c$ has been plotted instead of the resonance field H. Similar results have been given by Lax, Zeiger and Dexter (1954).

Table 4.1.

	Electrons		Holes				
Substance	m_1	m_2	A	$\lvert B\rvert$	$\lvert C\rvert$	Ref.	Remarks
Germanium	0·082	1·64	−13·1	8·3	12·5	a	(111)
	0·082	1·58	−13·2	8·9	10·6	b	
Silicon	0·19	0·98	− 4·0	1·1	4·1	a	(100)
	0·19	0·97	− 4·1	1·6	3·3	b	
Indium antimonide	0·02					c	
	0·013					d	

m_1, m_2 in units of m; A, B, C in units of $\hbar^2/2m$; (111), (100): electron energy surfaces spheroids oriented along (111, 100) directions of Brillouin zone.

References: a, Dexter, Zeiger and Lax 1956; b, Dresselhaus, Kip and Kittel 1955 a; c, Dexter and Lax 1955 b; d, Dresselhaus, Kip and Kittel 1955 b.

The corresponding analysis for holes at the top of the valence band is a difficult mathematical problem. Approximate solutions (Dresselhaus, Kip and Kittel 1955 a, Lax, Zeiger and Dexter 1954) have been given for the angular variation of the effective mass parameter when the magnetic field H lies in a (110) plane. If the angle θ is measured from the (100) axis

$$m^* = \frac{\hbar^2}{2}\frac{1}{A \pm [B^2+(\frac{1}{2}C)^2]^{1/2}}\left\{1 \pm \frac{C^2(1-3\cos^2\theta)^2}{64[B^2+(\frac{1}{2}C)^2]^{1/2}\{A \pm [B^2+(\frac{1}{2}C)^2]^{1/2}\}}\cdots\right\}. \qquad \ldots\ldots(4.11)$$

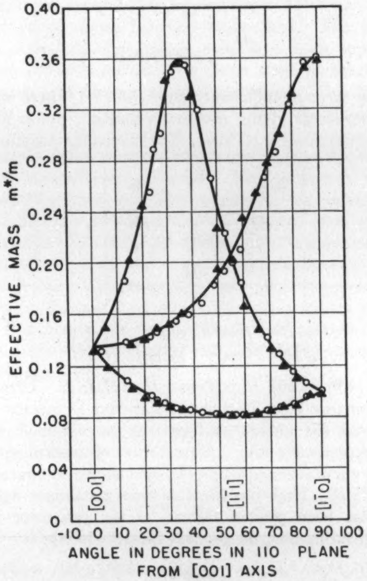


Figure 4.2. Effective masses of electrons in germanium at 4°K for magnetic field directions in a (110) plane (*after* Dresselhaus *et al.* 1955 a).

In the case of germanium and silicon, two lines are usually observed (figure 4.3) and by fitting the angular variation, the best values of A, B, C have been derived (see table 4.1).

Some preliminary measurements have been reported for indium antimonide (see table 4.1) (Dresselhaus *et al.* 1955 b, Dexter and Lax 1955 b), and for germanium–silicon alloys (Dresselhaus, Kip, Han-Ying Ku *et al.* 1955). The effective mass parameters and orientation of the energy surfaces differ considerably in germanium from silicon, although each has the diamond type lattice structure.

Measurements on alloys are therefore of considerable importance but the results reported so far are not sufficiently conclusive to provide any definite information about the changes in effective mass parameters upon alloying.

The relaxation time which characterizes the rate at which a carrier can exchange energy with the lattice can be derived from the line width. Its value is of the same order for all semiconductors investigated so far and is about 5×10^{-11} sec at 4°K but Fletcher *et al.* (1955) have reported one extremely pure crystal of germanium with a collision time of 3×10^{-10} sec at 1·3°K. The phonon part of the interaction should be temperature dependent. The measurements on germanium and silicon indicate that even in these relatively pure materials the major part of the relaxation process arises from interactions with impurity sites and lattice defects.

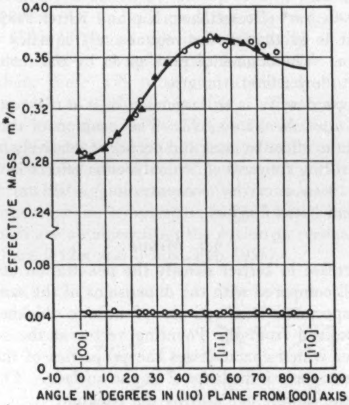


Figure 4.3. Effective masses of holes in germanium at 4°K for magnetic field directions in a (110) plane (*after* Dresselhaus *et al.* 1955 a).

Cyclotron resonance at infra-red wavelengths ($\lambda \sim 10\mu, 40\mu$) and 290°K has been reported for InSb (Burstein, Picus and Gebbie 1956, Zwerdling *et al.* 1956). Such high frequencies are required at room temperature in order to satisfy the condition $\omega/\nu > 1$.

4.4. *Plasma Effects*

When the electron density increases the microwave field acting on the charge carriers is modified in the first instance by space charge effects and later by the skin depth penetration. The space charge contribution can be written in equation (4.3) as a term proportional to the polarization. The skin depth is still assumed to be large compared with the sample dimensions, and the polarization is of the form

$$\mathbf{P} = Ne\mathbf{r} + \chi(\mathbf{E} - \alpha\mathbf{P}). \qquad \ldots\ldots(4.12)$$

The first term arises from the displacement of the ' free ' electrons with respect to the positive ion cores and the second term arises from the crystal lattice itself. For an isotropic effective mass, and a time variation $\exp(j\omega t)$ the equation of motion derived from equation (4.3) is

$$\hbar\dot{\mathbf{k}} = -e\left[(\mathbf{E}-\alpha\mathbf{P}) + \frac{1}{c}(\mathbf{v}\times\mathbf{H})\right] - \nu\hbar\mathbf{k} \qquad \ldots\ldots(4.13)$$

giving

$$\left[j\left(\omega - \frac{\omega_p{}^2}{\omega}\right) + \frac{\nu}{m^*}\right]\mathbf{v} = \frac{-e\mathbf{E}}{(1+\alpha\chi)\,m^*} - \frac{e}{m^*c}(\mathbf{v}\times\mathbf{H}) \qquad \ldots\ldots(4.14)$$

with $\omega_p{}^2 = \alpha Ne^2/(1+\alpha\chi)\,m^*$ (Dresselhaus, Kip and Kittel 1955 b). Evidently the resonance behaviour is as though the microwave frequency were reduced to a value $\omega' = \omega - \omega_p{}^2/\omega$. The resonance peak given by the relation $\omega' = eH/m^*c$ is therefore displaced to lower field strengths.

When the electron density is so large that $\omega_p \gg \omega$ the resonance condition is $\omega_p{}^2/\omega = eH/m^*c$ or $\omega = \alpha Nec/(1+\alpha\chi)H$. The position of the resonance peak is now independent of the effective mass and occurs at relatively lower field strengths the greater the measuring frequency. Such plasma effects have been reported in indium antimonide with a carrier concentration $\sim 10^{14}\ \mathrm{cm}^{-3}$ at 4°K and 77°K (Dresselhaus, Kip and Kittel 1955 b).

4.5. *Metals*

With further increase in carrier density the penetration depth for the microwaves becomes small compared with the dimensions of the sample. Under these circumstances the specimen is effectively of infinite thickness and the power absorbed may be derived from the Poynting vector at the surface. The most convenient parameter which characterizes the properties of the specimens under these conditions is the wave impedance Z at the surface. This is related to the power flowing into the sample per unit by the equation

$$P = \left(\frac{c}{4\pi}\right)^2 |\mathbf{H}_m|^2 \mathscr{R}(Z). \qquad \ldots\ldots(4.15)$$

Here $\mathbf{H}_m$ is the microwave magnetic field immediately outside the surface and Z is given by $Z = (j\omega\mu/4\pi\sigma)^{1/2}$ with μ the permeability and σ the conductivity of the specimen which in this case is complex, $\sigma = \sigma_1 - j\sigma_2$ (compare § 3). We then obtain

$$Z = \left(\frac{\omega\mu}{8\pi|\sigma|^2}\right)^{1/2} \{(|\sigma| - \sigma_2)^{1/2} + j(|\sigma| + \sigma_2)^{1/2}\}. \qquad \ldots\ldots(4.16)$$

The conductivity σ can be calculated when the energy surfaces are known. If the effective mass is isotropic it seems that there may be no resonant absorption lines of the type observed in semiconductors with low densities of carriers, but for more complicated Fermi surfaces singularities in the high-frequency conductivity may occur. Measurements have been reported for single crystals of bismuth (Dexter and Lax 1955 a, Galt *et al.* 1955) and theoretical discussions have been given by Tinkham (1956) and by Lax *et al.* (1956). Very recently Fawcett (1956) has reported resistance anomalies in single crystals of tin and copper. In this case the steady magnetic field was applied in the plane of the surface of the specimen. Such an effect had been predicted by Azbel and Kaner (1956).

4.6. *Non-resonant Absorption in Semiconductors*

The Drude classical theory of free charge carriers has been used by several authors to describe the propagation of an electromagnetic wave through a semiconductor when the resistivity is sufficiently high. If the microwave period is long compared with the collision time τ of the carriers ($\omega\tau \ll 1$), the propagation constant K is given by $K = 2\pi\kappa/\lambda = (4\pi/c)(\sigma/\sqrt{\epsilon})$, where κ is the imaginary part of the complex refractive index, σ is the d.c. conductivity, ϵ is the dielectric constant and c is the velocity of light in free space (Fan and Becker 1951). This relation has been verified at 1·25 cm by Gibson (1956), who assumed that $\sqrt{\epsilon}$ had a constant value equal to 4·1 and assigned the observed change in attenuation between 15° C and 55° C to the change in conductivity of the sample. If the collision time for the carriers is sufficiently long for $\omega\tau$ to be an appreciable fraction of one radian, there will be a contribution from the carriers to the dielectric constant as well as to the conductivity. The temperature variation of the dielectric constant may then be used to obtain an estimate of the collision time and, combining this with the mobility, a value for a mean effective mass of the carriers (Benedict and Shockley 1953, Benedict 1953, D'Altroy and Fan 1954). These experiments suggest a mean effective mass of about $0{\cdot}3m$ for holes and $0{\cdot}6m$ for electrons over the temperature range 150° K to 300° K. The corresponding values which are obtained from the cyclotron resonance results at lower temperatures (§ 4.3) are $0{\cdot}12m$ and $0{\cdot}16m$ (Fan 1955).

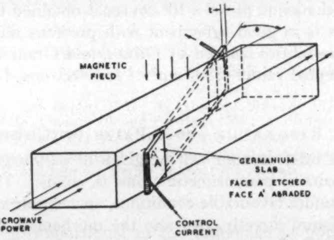


Figure 4.4. Diagram of a microwave attenuator using germanium (*after* Gunn and Hogarth 1955).

Gunn and Hogarth (1955) have made use of the dependence of absorption coefficient upon conductivity and devised a microwave attenuator and modulator (figure 4.4). One of the broad faces of a slab of long-lifetime germanium was abraded and the other etched, thus giving large and small values of the recombination velocity s at opposite surfaces of the slab. Ohmic connections were made to the ends and a magnetic field was applied in the plane of the broad faces and perpendicular to the direction of current flow. This field deflected both holes and electrons (they travel in opposite directions) either towards the surface of large s or away from it depending on the direction of the current. In the first case the rapid recombination at the surface decreased the carrier concentration and also decreased the attenuation for microwaves. In the second case free carrier pairs generated at the surface recombined only slowly and the conductivity and attenuation were increased. The minimum attenuation for this design is given as 3·7 dB whilst the maximum attenuation was 17·7 dB (for d.c. control current 10 mA) or 33 dB (for pulsed control current 60 mA). The response time was of the order of 1 μsec.

The non-resonant absorption of microwave radiation by semiconductors has also been applied to the measurement of the drift mobility of minority carriers in germanium (Gibson and Granville 1956). A filament of germanium (6 mm × 2 mm × 0·3 mm) was inserted through a slot cut in the broad face of 8 mm waveguide, the narrow dimension of the guide having been reduced to 1·5 mm in order to increase the resolving power of this method of measurement. The minority carriers were liberated at a point in the sample a few millimetres from the waveguide by optical excitation, and they were swept towards the guide by a rectangular voltage pulse. Their arrival within the waveguide was detected by the additional microwave absorption produced by the increased carrier concentration. If the distance of the light spot on the germanium from the waveguide is x, and the time interval between the leading edge of the voltage pulse and arrival of the carriers at the waveguide is Δt, the mobility μ is given by $x = \mu E \Delta t$. The amplitude of the voltage pulse was varied to provide field strengths E up to 5000 vcm^{-1}. The measurements were slightly complicated by an initial decrease in absorption due to the change in the mobility of the residual carriers at high field strengths (Arthur *et al.* 1956) but it was possible to verify that the drift velocity of electrons in p-type germanium saturated at high electric field strengths and the value of $5{\cdot}8 \times 10^6$ cm sec^{-1} obtained for the saturated drift velocity of electrons is in good agreement with previous measurements. At low field strengths the mobilities derived by Gibson and Granville were in agreement with the usually accepted values (3800 $\mathrm{cm}^2\mathrm{v}^{-1}$ for electrons, 1800 $\mathrm{cm}^2\mathrm{v}^{-1}$ for holes).

§ 5. SPIN RESONANCE FROM PAIRS OF COUPLED ATOMS

In this section a brief account will be given of paramagnetic resonance from exchange coupled pairs of paramagnetic atoms in solids. The measurements are of interest because, under favourable conditions, an accurate value for the exchange interaction can be found directly, and also the mechanism of the interaction can be studied more easily than for complicated systems. Some examples are in table 5.1.

For all of these examples the two magnetic atoms each have one unpaired electron, and the same effective electronic spin $s' = s'' = \frac{1}{2}$. It has been found (see references in table 5.1) that the main features of the paramagnetic resonance spectrum from the pair can be formally accounted for by assuming an exchange interaction of the isotropic type

$$\mathscr{H} = J\mathbf{s}'.\mathbf{s}''. \qquad \ldots\ldots(5.1)$$

This may be written

$$\mathscr{H} = \tfrac{1}{2}J[(\mathbf{s}'+\mathbf{s}'')^1 - \mathbf{s}'^2 - \mathbf{s}''^2$$

and the system can be described by a total spin vector $\mathbf{S} = \mathbf{s}' + \mathbf{s}''$. For the present case $s' = s'' = \frac{1}{2}$, so that

$$J\mathbf{s}'.\mathbf{s}'' = \tfrac{1}{2}J[S(S+1) - \tfrac{3}{2}] \qquad \ldots\ldots(5.2)$$

where $S = 1$ or 0, and the energy levels are a triplet ($S = 1$) at $J/4$, and a singlet ($S = 0$) at $-3J/4$. In a spin resonance experiment transitions can normally be induced only between the Zeeman components of the triplet ($\Delta S_z = \pm 1$). The magnitude of J can then be estimated (in principle) from the pair spectrum in several ways, for example, (*a*) temperature dependence of signal intensity, (*b*) resolution of lines with different *g*-values, (*c*) effect on hyperfine structure. In addition, in some cases there is an anisotropic interaction between the spins which gives a zero field splitting of the triplet levels, and this can also be measured. We discuss these four effects in turn.

Table 5.1. Some Examples of Exchange Coupled Pairs of Paramagnetic Ions in Crystals

Salt and paramagnetic ion	Approximate structure of pair	J (cm^{-1})	Observed effect in spectrum	Reference
$\mathrm{Cu(CH_3COO)_2H_2O}$ $\mathrm{Cu^{2+}}$, $3\mathrm{d}^9$, $s = \frac{1}{2}$	Cu–Cu sepn. $\simeq 2{\cdot}6$ Å	$+300$	$S = 1$ hyperfine structure intensity	a, b, c
$\mathrm{K_2(Ir,Pt)Cl_6}$ 1 : 20 $\mathrm{Ir^{4+}}$, $5\mathrm{d}^5$, $s = \frac{1}{2}$	$\mathrm{Ir}\left\langle\begin{matrix}\mathrm{Cl{-}Cl}\\ \mathrm{Cl{-}Cl}\end{matrix}\right\rangle\mathrm{Ir}$ sepn. $\simeq 6{\cdot}9$ Å	$+7$	$S = 1$ intensity	d
As in Si 1 : 100000 As, $4\mathrm{s}^2 4\mathrm{p}^3$, $s = \frac{1}{2}$	As ... As sepn. = ?	$\lvert J \rvert$ $\gg 0{\cdot}01$	hyperfine structure	e, f
$\mathrm{Cu(SO_4)5H_2O}$ $\mathrm{Cu^{2+}}$, $3\mathrm{d}^9$, $s = \frac{1}{2}$	(Cu–O–O–Cu)† sepn. $\simeq 5{\cdot}2$ Å	$\lvert J \rvert$ $\sim 0{\cdot}15$	average *g*-value	g, h

References: a, Bleaney and Bowers 1952; b, Abe and Shimada 1953; c, van Niekerk and Schoening 1953; d, Griffiths, Owen, Park and Partridge 1957; e, Slichter 1955; f, Fletcher *et al.* 1954; g, Bagguley and Griffiths 1948; h, Pryce 1948.

† Each Cu has six fairly close Cu neighbours at the corners of a distorted octahedron; of these, two are dissimilar at 5·2 Å, and four are similar at about 6 Å distance.

(*a*) *Signal intensity.*

The intensity R is proportional to the difference in population between the two spin levels (separation $g\beta H$) between which transitions occur. For the normal case of a ground state with effective spin $s = \frac{1}{2}$, well separated from higher states, $R \propto g\beta H/kT$ if $g\beta H \ll kT$. For the present case where resonance is only observed between the levels of a triplet, separated by J from a singlet, the corresponding result is

$$R \propto \frac{g\beta H}{kT}\,\frac{e^{-J/kT}}{1+3e^{-J/kT}}. \qquad \ldots\ldots(5.3)$$

If J is positive (antiferromagnetic) so that the singlet is lowest, and if $g\beta H \ll J$, there is a sharp drop in intensity if the temperature can be lowered to give $kT \ll J$.

An example is copper acetate, where Bleaney and Bowers (1952) find the signal intensity to be consistent with $J \simeq +300\ \mathrm{cm}^{-1}$. For this particular salt, all of the Cu ions occur in pairs (and behave rather like $(\mathrm{Cu^{2+}})_2$ molecules), so the bulk susceptibility also follows equation (5.3) (Guha 1951). A second example is $K_2(Ir, Pt)Cl_6$ where Griffiths, Owen, Park and Partridge (1957) find the signal intensity from nearest neighbour pairs of Ir ions to be consistent with $J \simeq +7\ \mathrm{cm}^{-1}$.

(*b*) *Averaging of g-values.*

We consider the case of two paramagnetic atoms, with effective spins $s' = s'' = \frac{1}{2}$, and isotropic but different g-values $g' \neq g''$. With applied magnetic field H, the spin-Hamiltonian can be written

$$\mathscr{H} = g'\beta\mathbf{H}.\mathbf{s}' + g''\beta\mathbf{H}.\mathbf{s}'' + J\mathbf{s}'.\mathbf{s}''. \qquad \ldots\ldots(5.4)$$

If $J = 0$ (or $\ll (g'-g'')\beta H$ and the line width) there are then two distinct resonance lines corresponding to frequencies $h\nu' = g'\beta H$ and $h\nu'' = g''\beta H$.

If $J \neq 0$, it is convenient to rearrange equation (5.4) (Pryce 1948), and to use equation (5.2), giving

$$\mathscr{H} = \tfrac{1}{2}(g'+g'')\beta\mathbf{H}.\mathbf{S} + \tfrac{1}{2}J[S(S+1)-\tfrac{3}{2}] + \tfrac{1}{2}(g'-g'')\beta\mathbf{H}.(\mathbf{s}'-\mathbf{s}'') \qquad \ldots\ldots(5.5)$$

where $\mathbf{S} = \mathbf{s}' + \mathbf{s}''$ and $S = 1$ or 0. The effect of the last term is to mix the centre level of the triplet ($S_z = 0$ belonging to $S = 1$) with the singlet $S = 0$. It can be neglected if it is small compared with the separation between these levels, i.e. if $(g'-g'')\beta H \ll J$. The triplet energy levels are then $\frac{1}{4}J$ and $\frac{1}{4}J \pm \frac{1}{2}(g'+g'')\beta H$, and there is only one transition frequency $h\nu = \frac{1}{2}(g'+g'')\beta H$ corresponding to the mean g-value. For the intermediate case of $(g'-g'')\beta H \simeq J$ one might expect to find four transition frequencies.

A salt which illustrates the effect is copper sulphate. The general theory has been given by Pryce (1948) (see also Yosida 1950). Here there are not isolated pairs of Cu ions, but each ion interacts to a greater or lesser extent with several neighbours (see table 5.1) which influences the line width (see Pryce 1948). The important point for the present purpose is that the two types of ion in unit cell are fairly strongly coupled by exchange, and for most directions in the crystal they have different g-values. Bagguley and Griffiths (1948) find that for a particular direction, at a microwave frequency $1{\cdot}2\ \mathrm{cm}^{-1}$, there are two well-resolved lines with $g' = 2{\cdot}46$, $g'' = 2{\cdot}08$. On the other hand, with frequency $0{\cdot}3\ \mathrm{cm}^{-1}$ there is only one line corresponding to the mean g-value, $g = 2{\cdot}27$. For the intermediate case of frequency $0{\cdot}8\ \mathrm{cm}^{-1}$ the lines are almost resolved. This last case corresponds to $(g'-g'')\beta H \sim |J|$, giving $|J| \sim 0{\cdot}15\ \mathrm{cm}^{-1}$ for the exchange interaction between dissimilar ions. This is the result proved in a more general way by Pryce. Similar effects have been found in $K_2CuCl_4 2H_2O$ by Abe *et al.* (1954). For salts where the exchange is expected to be about ten or more times larger (e.g. $CuCl_2 2H_2O$, § 9) only one line has been found in the paramagnetic resonance spectrum at normal microwave energies $\sim 1\ \mathrm{cm}^{-1}$.

(*c*) *Effects on hyperfine structure.*

This case has been considered by Slichter (1955) (see also Bleaney and Bowers 1952). We assume $g' = g''$ and that each unpaired electron is coupled to its nucleus of spin I ($I' = I''$) with hyperfine interaction A. The Hamiltonian is then

$$\mathscr{H} = (g\beta\mathbf{H}.\mathbf{s}' + A\mathbf{s}'.\mathbf{I}') + (g\beta\mathbf{H}.\mathbf{s}'' + A\mathbf{s}''.\mathbf{I}'') + J\mathbf{s}'.\mathbf{s}''. \qquad \ldots\ldots(5.6)$$

If $J = 0$ and if $g\beta H \gg A$, the energy levels are $\pm\frac{1}{2}g\beta H \pm \frac{1}{2}AI_z$ so that the spectrum contains $2I'+1 = 2I''+1$ hyperfine structure lines with magnetic field spacing $A/g\beta$ between adjacent lines.

If $J \neq 0$, it is convenient (cf. eqn (5.5)) to write (5.6)

$$\mathcal{H} = g\beta\mathbf{H}\,.\,\mathbf{S} + \tfrac{1}{2}A\mathbf{S}\,.(\mathbf{I}'+\mathbf{I}'') + \tfrac{1}{2}J[S(S+1)-3/2] + \tfrac{1}{2}A(\mathbf{s}'-\mathbf{s}'').(\mathbf{I}'-\mathbf{I}'') \qquad \ldots\ldots(5.7)$$

with the total spin $S = 1$ or 0. The last term can be neglected if $J \gg A$. Then, if $g\beta H \gg A$, the energy levels belonging to $S = 1$ are $\frac{1}{4}J$, $\frac{1}{4}J \pm g\beta H \pm \frac{1}{2}A(I_z'+I_z'')$. The spectrum from transitions $\Delta S_z = \pm 1$, $\Delta I_z = 0$, then contains $2(I'+I'')+1$ hyperfine structure lines, with spacing $\frac{1}{2}A/g\beta$ between adjacent lines, and with relative intensities corresponding to the number of ways of forming each value of $I_z'+I_z''$.

An example which illustrates this intensity relation is copper acetate, where $I = 3/2$ for ^{63}Cu and ^{65}Cu, the two isotopes having nearly equal magnetic moments. Bleaney and Bowers (1952) find the relative intensities of the hyperfine lines in the pair spectrum to be 1, 2, 3, 4, 3, 2, 1. An example which shows that the separation of the hyperfine lines from a pair is only half of that from a single atom, is the donor impurity As in Si ($\sim 10^{17}$ As atoms/cm^3), discussed in § 2.2. The paramagnetic resonance spectrum (Slichter 1955, Fletcher *et al.* 1954) shows that most of the As atoms are non-interacting giving four equally spaced, equally intense, hyperfine lines corresponding to $I = 3/2$ for ^{75}As. There are some coupled pairs, however, which give a weaker spectrum of lines with half the spacing and relative intensities 1, 2, 3, 4, 3, 2, 1. The 1, 3, 3, 1 lines coincide with the strong single atom lines and are not seen. The 2, 4, 2 lines occur midway between the strong lines as observed. Groups of three and four exchange coupled donors have also been detected (Feher, Fletcher and Gere 1955). It will be noted that if the concentration is high enough for a large number of atoms to be interacting simultaneously, the effect of exchange is to give many closely spaced (unresolved) hyperfine lines with high intensity in the centre. There is then a single exchange narrowed line (cf. Anderson 1954, van Wieringen 1955).

Up to the present time there appear to be no definite experimental examples of the intermediate case where $J \sim A$, although this might happen for magnetic centres in irradiated Perspex (Schneider 1955, § 2.4).

(*d*) *Anisotropic coupling*

So far, we have considered experimental results which are consistent with an isotropic spin–spin coupling, $J\mathbf{s}'.\mathbf{s}''$. Here the interaction energy depends only on the orientation of one spin relative to the other. In some cases, however, it is found experimentally that the interaction energy is also dependent on the orientation of the coupled spins relative to their surroundings in the crystal (which implies that there is a zero field splitting of the triplet levels belonging to $S = 1$). This is represented by introducing an anisotropic spin–spin coupling, which is usually said to be the result of 'anisotropic exchange'. Van Vleck (1937) has suggested that for the case of two spins with $s = \frac{1}{2}$, a suitable anisotropic interaction would result if the spins were coupled by isotropic exchange and in addition each spin were coupled to its surroundings (of low symmetry) in the crystal via spin–orbit and orbit–crystal field interactions (see also Kramer, 1934, Opechowski 1948, Bleaney and Bowers 1952, Stevens 1953).

If $s' = s'' = \frac{1}{2}$, an interaction with rhombic symmetry can be formally represented (Kramers 1930, Van Vleck 1937, Baker and Bleaney 1955) by adding terms to the Hamiltonian†

$$\begin{aligned}\mathscr{H}_a &= D(2s_z' s_z'' - s_x' s_x'' - s_y' s_y'') + E(s_x' s_x'' - s_y' s_y'') \\ &= \tfrac{3}{2}D(S_z^2 - \tfrac{2}{3}) + \tfrac{1}{2}E(S_x^2 - S_y^2) \qquad \ldots\ldots(5.8)\end{aligned}$$

where S is the total spin with values 1 or 0. The separation J between the singlet ($S = 0$) and the centre of gravity of the triplet ($S = 1$) is unchanged by this interaction, but the triplet levels are now at $-D, \frac{1}{2}D \pm \frac{1}{2}E$ relative to their centre of gravity. If the symmetry is uniaxial, as would happen if the line joining the two ions was an axis of symmetry both of the ions and their surroundings, then $E = 0$, (5.8) has the form of a dipole–dipole interaction, and the triplet levels are split into a doublet and a singlet separated by $\frac{3}{2}D$. It may be remarked that the true magnetic dipolar interaction is included by the interaction (5.8), but in many cases it represents only a small part of the total interaction.

An example with approximately axial symmetry is copper acetate, where Bleaney and Bowers find the splitting $\frac{3}{2}|D| = 0{\cdot}34$ cm^{-1}. This compares with the isotropic exchange, $J = 300$ cm^{-1}. These authors were able to find a relation between J and D in terms of the spin–orbit coupling and the crystal field, which was in fair agreement with experiment (see also Stevens 1953). An example where there is lower than axial symmetry is the interaction between nearest neighbour Ir ions in $K_2(Ir, Pt)Cl_6$, where Griffiths and Partridge find

$$\tfrac{3}{2}|D| = 0{\cdot}675 \text{ cm}^{-1}, \ \tfrac{1}{2}|E| = 0{\cdot}089 \text{ cm}^{-1},$$

and D/E is positive. These values compare with the isotropic exchange $J = +7$ cm^{-1}.

From the microwave point of view, one advantage of having anisotropic coupling and a zero field splitting of the levels is that the spectrum shows unambiguously that the coupled system has total spin $S = 1$, and also, for the case of dilute crystals, the symmetry properties of the spectrum suggest which neighbours are interacting.

(*e*) *Concentrated crystals*

In concentrated crystals each paramagnetic atom normally interacts with many neighbours. The effects of isotropic and anisotropic spin–spin interactions then include the following : (i) there is a contribution to the specific heat, and to the Weiss constant in the susceptibility (see, for example, Ollom and Van Vleck 1951) ; (ii) isotropic exchange may cause a transition to antiferromagnetism (J positive) or possibly ferromagnetism (J negative), and in both cases anisotropic exchange may contribute to the anisotropy energy ; (iii) isotropic exchange may, under certain conditions, lead to a narrowing of the paramagnetic resonance line, and anisotropic exchange to a broadening (see, for example, Van Vleck 1948,

---

† This is sometimes written in the form

$$\mathscr{H}_a = \mathbf{S}'.\mathbf{J}_a.\mathbf{S}'' \equiv \mathcal{J}_x S_x' S_x'' + \mathcal{J}_y S_y' S_y'' + \mathcal{J}_z S_z' S_z''$$

where $\mathcal{J}_x + \mathcal{J}_y + \mathcal{J}_z = 0$ and $D = \frac{1}{2}\mathcal{J}_z$, $E = \frac{1}{2}(\mathcal{J}_x - \mathcal{J}_y)$.

Pryce 1948, Bleaney and Stevens 1953, Anderson and Weiss 1953, Baker and Bleaney 1955). Also, for a general symposium on exchange interactions, see *Reviews of Modern Physics*, 1953, **25**, 191. For the present purpose we note that exchange interactions can sometimes be estimated from the line width but normally this method can only deal with interactions $\sim 0{\cdot}1\ \mathrm{cm}^{-1}$ or less, which is of the order of the measurable line width.

One of the interests in measuring pair spectra in dilute crystals is that exchange energies of many cm^{-1} (such as occur in normal antiferromagnetics) can be found directly. The values can then be assumed to apply for concentrated crystals provided the crystal structure is unchanged. Not much work along these lines has yet been done. One example is $K_2(Ir, Pt)Cl_6$ where the pair spectrum gives $J = +7\ \mathrm{cm}^{-1}$ as explained above. In the concentrated salt, each Ir ion at sites such as $(0, 0, 0)$ has twelve nearest neighbour Ir ions at sites such as $(\frac{1}{2}, \frac{1}{2}, 0)$ in the face-centred cubic structure. One might then expect (cf. Van Vleck 1941, Anderson 1950) the high temperature susceptibility to follow a Curie–Weiss law $\chi = C/(T+\theta)$ with $\theta = 3J/k \simeq 30°\mathrm{K}$. This is not inconsistent with preliminary measurements of Cooke, McKim, Smith and Wolf (1957). The exchange may also lead to an antiferromagnetic transition and a specific heat anomaly which is found by these authors to be at about $3{\cdot}3°\mathrm{K}$.

§ 6. GENERAL DISCUSSION OF EXTENSIVE COUPLED SYSTEMS

It is not possible to extend the detailed microscopic treatment which has been given for pairs of spins to the many-body problem which arises in connection with ferromagnetic, ferrimagnetic and antiferromagnetic materials. In these cases there are of the order 10^{22} interacting spins and it is then convenient to treat the problem in terms of the macroscopic magnetization $\mathbf{M}$. The magnetization $\mathbf{M}$ is the vector sum of the moments $\mathbf{m}_i$ of the individual carriers and these are written in terms of the spin only angular momentum $\mathbf{s}_i\hbar$; thus $\mathbf{m}_i = (e\hbar/2mc)g_i\mathbf{s}_i$ and $\mathbf{M} = \Sigma\mathbf{m}_i = (e\hbar/2mc)\Sigma g_i\mathbf{s}_i$. This relation defines g_i to be the spectroscopic splitting factor as used in paramagnetic resonance. When g_i is different from $2{\cdot}0$ the magnitude $|g_i - 2|$ is a measure of the spin–orbit interaction. The significance of this definition is discussed in the quantum mechanical treatments of Kittel (1949), Polder (1949) and Van Vleck (1950). These authors show that, provided the spin–orbit coupling is weak, the spin angular momentum in the direction of the resultant magnetic field $\mathbf{H}'$ which is acting upon the system changes by $\hbar$ for first order transitions. An absorption line then occurs at a frequency given by the relation

$$\hbar\omega_r = \frac{e\hbar}{2mc}gH' = g\beta H'. \qquad \ldots\ldots(6.1)$$

It is to be noted that in the gyromechanical experiments, the gyromagnetic factor g' is defined in terms of the total angular momentum of the system ($\mathbf{L} + \mathbf{S} = \mathbf{J}$). This difference in definition leads to the relation $g - 2 = 2 - g'$ (Kittel 1949, Van Vleck 1950). It is possible to show this in an approximate way for simple systems where all the g_i are equal. In the case of magnetic resonance, the magnetization is written in terms of the spin only angular momentum

$\langle \mathbf{L}+2\mathbf{S}\rangle\beta = g\langle \mathbf{S}\rangle\beta$ and so when $\mathbf{L}$ is practically quenched

$$g \sim \frac{\langle \mathbf{L}+2\mathbf{S}\rangle}{\langle \mathbf{S}\rangle} \sim 2 + \left\langle \frac{\mathbf{L}}{\mathbf{S}} \right\rangle .$$

For gyromagnetic experiments on the other hand, $\langle \mathbf{L}+2\mathbf{S}\rangle\beta = g'\langle \mathbf{L}+\mathbf{S}\rangle\beta$ and so when $\mathbf{L}$ is small

$$g' \sim \frac{\langle \mathbf{L}+2\mathbf{S}\rangle}{\langle \mathbf{L}+\mathbf{S}\rangle} \sim 2 - \left\langle \frac{\mathbf{L}}{\mathbf{S}} \right\rangle \quad \text{and so} \quad g-2 \sim \left\langle \frac{\mathbf{L}}{\mathbf{S}} \right\rangle \sim 2-g'.$$

The macroscopic magnetization is the vector sum of the individual contributions and so, in the case of identical carriers, the spectroscopic splitting factor derived from equation (6.1) is that which describes each individual spin. In many systems, however, for example ferrimagnetic materials or some ferromagnetic alloys, the magnetic carriers may not all have the same *g*-values. But even in these cases when the coupling between the different component systems is sufficiently large and anisotropy terms are neglected, there is only one precessional mode (Smit and Wijn 1954, Wangsness 1953). The magnetization can then be written

$$\mathbf{M} = \frac{e\hbar}{2mc}\Sigma g_i \mathbf{s}_i = \frac{e\hbar}{2mc} g_{\text{eff}} \Sigma \mathbf{s}_i$$

and the absorption line occurs at a position corresponding to

$$g_{\text{eff}} = \frac{\Sigma g_i \mathbf{s}_i}{\Sigma \mathbf{s}_i} = \frac{\Sigma \mathbf{m}_i}{\Sigma \mathbf{m}_i/g_i}.$$

We note that in the particular case of an antiferromagnetic material for which $\Sigma \mathbf{m}_i = 0$ and all the g_i are equal, this relation implies that the absorption line will not be displaced on cooling through the Curie point (provided that the anisotropy terms may be neglected).

When a magnetic system is placed in a magnetic field, it will experience a torque $\mathbf{M} \times \mathbf{H}$ which will produce a rate of change of angular momentum and consequently a rate of change of magnetization $(d\mathbf{M}/dt)(2mc/ge) = \mathbf{M} \times \mathbf{H}$. The actual value of the torque acting within an elementary volume of the material will be derived not only from the external applied field but also from fields due to the magnetization of the surrounding regions and also from anisotropy effects. Following Néel (1948) the field $\mathbf{H}_i$ acting on the ith system which arises from the surrounding regions, is written in the form $\mathbf{H}_i = -\Sigma_j \mathbf{N}_{ij}.\mathbf{M}_j$. Here $\mathbf{N}_{ij}$ is a tensor quantity. The torque arising from the anisotropy energy is written straightforwardly as $\Gamma_i = (\partial W_{\text{A}}/\partial\theta)_i$. A system described by such an equation of motion, however, is not magnetized by the applied field. The magnetic moment only precesses about the field direction and the equation needs to be modified to include a 'relaxation' term which enables the spin system to achieve thermal equilibrium with the lattice. Two suggestions have been put forward to describe this relaxation process. Landau and Lifshitz (1935) introduced a term

$$\mathbf{R} = -\lambda\left[\frac{\mathbf{M} \times (\mathbf{M} \times \mathbf{H})}{M^2}\right]$$

which is a vector quantity directed from $\mathbf{M}$ towards $\mathbf{H}$ and so may be regarded as representing the approach of $\mathbf{M}$ to its equilibrium direction. This term becomes zero when $\mathbf{M}$ lies along $\mathbf{H}$ and also gives rise to an energy dissipation as may be verified by evaluating $\int \mathbf{H}.(d\mathbf{M}/dt)$ so that it is somewhat analogous to a classical damping term. Bloembergen (1950) has used the Bloch (1946) equations and included the terms $R_{x,y} = -M_{x,y}/T_2$ for the x,y components and $R_z = -(M_z - M_0)/T_1$ for the z components of $d\mathbf{M}/dt$. Here T_1 and T_2 characterize an exponential decay of the components of magnetization towards their equilibrium values. The complete equation of motion for the magnetization in the ith system is written, therefore, as

$$\left(\frac{2mc}{e}\right)\frac{d\mathbf{M}_i}{dt} = g_i \mathbf{M}_i \times \left[\mathbf{H}^0 + \mathbf{h} - \sum_j \mathbf{N}_{ij}.\mathbf{M}_j\right] + g_i\left(\frac{\partial W_\mathrm{A}}{\partial\theta}\right)_i + \mathbf{R} \qquad \ldots\ldots(6.2)$$

where $\mathbf{H}^0$ is the external steady field and $\mathbf{h}$ is an applied time dependent field.

Fortunately a large number of physically significant features can be derived from approximate and simplified forms of equation (6.2). The anisotropy term can be handled by writing it as a vector product of a magnetization term and a fictitious anisotropy field such that $(\partial W_\mathrm{A}/\partial\theta) = \mathbf{M}^\mathrm{A} \times \mathbf{H}^\mathrm{A}$. The variable θ is a rotation about the direction of $\mathbf{M}^\mathrm{A} \times \mathbf{H}^\mathrm{A}$. This relation does not completely determine $\mathbf{M}^\mathrm{A}, \mathbf{H}^\mathrm{A}$ and allows the magnitude or direction of one term to be arbitrarily specified as most convenient for the particular experimental conditions. The time dependent field $\mathbf{h}$ is, in practice, very small compared with $\mathbf{H}^0$, and so $\mathbf{M}$ may be written $\mathbf{M}^0 + \mathbf{m}(t)$ with $\mathbf{M}^0$ parallel to $\mathbf{H}^0$. Neglecting products of small quantities $\mathbf{m}, \mathbf{h}$, the microwave behaviour can then be determined from the permeability which is obtained from the equation for the magnetization in the form

$$\begin{aligned}\left(\frac{2mc}{e}\right)\dot{\mathbf{m}}_i = g_i\Big[\mathbf{m}_i \times \mathbf{H}^0 + \mathbf{M}_i{}^0 \times \mathbf{h} - \sum_j \{\mathbf{M}_i{}^0 \times (\mathbf{N}_{ij}.\mathbf{m}_j) + \mathbf{m}_i \times (\mathbf{N}_{ij}.\mathbf{M}_j{}^0)\}\\ + \mathbf{M}_i{}^\mathrm{A} \times \mathbf{H}_i{}^\mathrm{A}\Big] + \mathbf{R}. \qquad \ldots\ldots(6.3)\end{aligned}$$

When all the constituent systems are identical, for example in the case of the ferromagnetic elements, the $\mathbf{m}_i$ are equal and the summation Σ_j in equation (6.3) represents the macroscopic demagnetizing factors N_x, N_y, N_z. This equation then becomes

$$\begin{aligned}\frac{2mc}{e}(\dot{\mathbf{m}}) = g\{\mathbf{m} \times \mathbf{H}^0 + \mathbf{M}^0 \times \mathbf{h} + [\mathbf{i}m_y(N_y - N_z)\,M^0 - \mathbf{j}m_x(N_x - N_z)\,M^0]\\ + \mathbf{M}^\mathrm{A} \times \mathbf{H}^\mathrm{A}\} + \mathbf{R}. \qquad \ldots\ldots(6.4)\end{aligned}$$

For ferrites or alloys which may have sub-lattices with different moments and g-factors the solution is approached by forming the sum

$$\sum_i \frac{2mc}{e}.\frac{\dot{\mathbf{m}}_i}{g_i} = \frac{2mc}{eg_\mathrm{eff}}\dot{\mathbf{m}}$$

over all the i systems. The double summation $\Sigma_i \Sigma_j$ which will occur on the right-hand side of equation (6·3) again provides effective demagnetizing factors but these terms are usually eliminated either by choosing a sample shape such that they will subtract out (for example, a sphere) or by discussing an infinite medium. Moreover, the anisotropy terms are usually small in the case of ferrites and may often be neglected when considering the underlying physical principles.

The equation of motion is then regarded as being the same as equation (6.4) except for the substitution of g_{eff} for the g-factor.

For resonance absorption the term $\mathbf{R}$ must be small and may be neglected in first approximation when deriving the resonant frequency. The anisotropy terms may be written $M^{\mathrm{A}} \equiv M^0$, $H_x{}^{\mathrm{A}} \equiv -N_x{}^{\mathrm{A}} m_x$, $H_y{}^{\mathrm{A}} \equiv -N_y{}^{\mathrm{A}} m_y$ with $N_x{}^{\mathrm{A}}, N_y{}^{\mathrm{A}}$ as effective demagnetizing factors (Kittel 1948), and taking $\mathbf{H}^0$ along the z axis we obtain from equation (6.4)

$$\begin{aligned} \left(\frac{2mc}{e}\right)\dot{m}_x &= g\{m_y[H^0+(N_y+N_y{}^{\mathrm{A}}-N_z)\,M^0]-M^0h_y\} \\ \left(\frac{2mc}{e}\right)\dot{m}_y &= g\{-m_x[H^0+(N_x+N_x{}^{\mathrm{A}}-N_z)M^0]+M^0h_x\} \\ \dot{m}_z &\simeq 0. \end{aligned} \qquad \ldots\ldots(6.5)$$

A solution of these equations with $\mathbf{m}$ of the form $\exp(j\omega t)$ is

$$\begin{aligned} m_x &= \frac{\gamma^2 H_1 M^0}{\gamma^2 H_1 H_2-\omega^2}h_x - \frac{j\omega\gamma M^0}{\gamma^2 H_1 H_2-\omega^2}h_y \equiv \mu_1 h_x - jkh_y \\ m_y &= \frac{j\omega\gamma M^0}{\gamma^2 H_1 H_2-\omega^2}h_x + \frac{\gamma^2 H_2 M^0}{\gamma^2 H_1 H_2-\omega^2}h_y \equiv jkh_x + \mu_2 h_y \\ M_z &= M^0 \equiv \mu_3 H^0. \end{aligned} \qquad \ldots\ldots(6.6)$$

Here $\quad H_1 \equiv H^0+(N_y+N_y{}^{\mathrm{A}}-N_z)\,M^0, \quad H_2 \equiv H^0+(N_x+N_x{}^{\mathrm{A}}-N_z)\,M^0$

$\gamma = ge/2mc$ and μ_1, μ_2, μ_3, k are components of a permeability tensor. The demagnetizing coefficients $N_x, N_x{}^{\mathrm{A}}, N_y, N_y{}^{\mathrm{A}}$ arise from the time dependent variables m_x, m_y whereas N_z derives from the steady magnetization M^0; thus they may only be expected to have their d.c. values when the effective dimensions of the sample are properly chosen. The imaginary terms in the expressions for m_x, m_y arise from the time differentials $\dot{h}_y, \dot{h}_x$. That they do not give rise to any power absorption may be verified by writing down the components and evaluating $\int h_x dm_x + h_y dm_y = 0$. This is to be expected since no relaxation terms were included in the equations of motion. The material is not isotropic, however, and there will be a rotation of the plane of polarization for an electromagnetic wave (see § 8.3).

An anomaly in the high frequency permeability occurs when

$$\omega = \gamma\sqrt{(H_1 H_2)} = (g\beta/\hbar)\sqrt{(H_1 H_2)},$$

this means that $\sqrt{(H_1 H_2)}$ corresponds to the effective field H' of equation (6.1).† Written out in full the resonance equation is

$$\hbar\omega_{\mathrm{r}} = g\beta\sqrt{\{H^0+(N_y+N_y{}^{\mathrm{A}}-N_z)\,M^0\}\{H^0+(N_x+N_x{}^{\mathrm{A}}-N_z)\,M^0\}}. \qquad \ldots\ldots(6.7)$$

This general expression has important particular cases. Writing ϕ_1, ϕ_2 for the anisotropy terms, the following relations are obtained for the different experimental arrangements (Kittel 1948, Van Vleck 1950):

(*a*) for H^0 and the microwave field in the plane of a plane sample

$$\hbar\omega_{\mathrm{r}} = g\beta\sqrt{\{H^0+\phi_1\}\{H^0+4\pi M^0+\phi_2\}}: \qquad \ldots\ldots(6.8)$$

---

† For a more rigorous identification of g and H' see Van Vleck (1950).

in this case the resonance line occurs at about 5000 Oe for nickel 1·25 cm wavelength ;

(*b*) for H^0 perpendicular to the sample and the microwave field in the plane of a plane sample

$$\hbar\omega_r = g\beta\sqrt{\{H^0-4\pi M^0+\phi_1\}\{H^0-4\pi M^0+\phi_2\}}: \qquad \ldots\ldots(6.9)$$

in this case the resonance line occurs at about 14 000 Oe for nickel at 1·25 cm wavelength ;

(*c*) for a spherical sample with dimensions small compared with the microwave wavelength in the material

$$\hbar\omega_r = g\beta\sqrt{\{H^0+\phi_1\}\{H^0+\phi_2\}}: \qquad \ldots\ldots(6.10)$$

in this case the resonance line occurs at about 8000 Oe for nickel 1·25 cm wavelength.

This latter equation has been widely used in connection with the measurements on single crystals of ferrites (see § 8). In the case of cubic crystals, the following relations in terms of the first and second order anisotropy constants K_1, K_2, have been given (Bickford 1950, Yager *et al.* 1950, Dillon *et al.* 1955) :

H^0 in a (100) direction $\hbar\omega_r = g\beta\{H^0+2K_1/M^0\}$,

H^0 in a (110) direction $\hbar\omega_r = g\beta\sqrt{\{H^0-2K_1/M^0\}\{H^0+K_1/M^0+\tfrac{1}{2}K_2/M^0\}}$,

H^0 in a (111) direction $\hbar\omega_r = g\beta\{H^0-4K_1/3M^0-4K_2/9M^0\}$. ......(6.11)

In the following sections of this report we shall review the microwave measurements which have been carried out with ferromagnetic, ferrimagnetic and antiferromagnetic materials. These experiments have usually been interpreted in terms of the general principles which we have set out above.

§ 7. FERROMAGNETIC METALS

The investigation of microwave resonance absorption in ferromagnetic metals has not been so spectacular a method for deriving detailed information about the magnetic carriers as has been the case for paramagnetic salts. There are a number of reasons for this, for example it is difficult to obtain good strain free single crystal surfaces, the line widths are relatively broad, and the strongly coupled system does not reveal any detailed structure in its absorption spectrum. Moreover, because of the large magnetization (~ 1000 gauss) the demagnetization and anisotropy terms in equation (6.7) may make important contributions to the resonance condition so that some care is necessary in order to devise an experimental arrangement from which the results may be fairly readily interpreted. We shall therefore give a brief outline of the experimental principles before reviewing the results which have been obtained and their theoretical interpretation.

7.1. *Experimental Principles*

Two methods of investigating ferromagnetic resonance phenomena in metals have been reported in the literature. The method most widely used (e.g. Griffiths 1946, Yager 1949, Bloembergen 1950, Standley and Reich 1955) measures the input impedance of a plane ferromagnetic surface which forms one wall of a cavity resonator. The other method (Bagguley 1953, 1955) measures the power absorbed in a dilute suspension of colloidal ferromagnetic particles, which are assumed to be spherical.

In the first method, the steady magnetic field and the microwave magnetic field are mutually perpendicular and lie in the plane of the surface of the specimen. The spectroscopic splitting factor is then derived from equation (6.8) provided that the line width is not too broad. In order to apply equation (6.8) it is necessary to know the demagnetizing factors, strain and anisotropy corrections, and the saturation magnetization appropriate to the surface layers of the sample (the ferromagnetic metal is a relatively good conductor and the skin depth is only of the order of 10^{-5} cm at 1 cm wavelength). These terms can make considerable contributions to the resonance conditions and are not necessarily obtained from d.c. measurements on bulk specimens (Griffiths 1951, Griffiths and MacDonald 1951). Thus Standley (1949) found that strains in nickel films could cause an apparent decrease in the g-factor for nickel from 2·20 (unstrained) to 2·05 (strained).

The surface impedance of the ferromagnetic sample may be written in terms of the complex permeability $\mu = \mu' - j\mu''$. The term μ'' which gives rise to the resonance power absorption was not included in equation (6.6) because in deriving that expression the relaxation term **R** was neglected in order to simplify the mathematics. The surface impedance is

$$Z = \left(\frac{j\omega\mu}{4\pi\sigma}\right)^{1/2} = \left(\frac{\omega}{8\pi\sigma}\right)^{1/2} \{(|\mu| + \mu'')^{1/2} + j(|\mu| - \mu'')^{1/2}\}. \qquad \ldots\ldots(7.1)$$

The power absorbed is proportional to $\mathscr{R}(Z)$ and so depends upon $(|\mu| + \mu'')^{1/2}$. In order to derive μ', μ'' separately, a further experiment must be carried out to determine $(|\mu| - \mu'')^{1/2}$ from the change in resonant frequency of the cavity within the absorption line. A typical absorption curve for $(|\mu| + \mu'')^{1/2}$ is shown in figure 7.1. There is a minimum absorption at a magnetic field given by the relation $\omega_{\min} = \gamma(H_{\min}{}^0 + 4\pi M^0)$ provided the line width is small, that anisotropy terms and variations of μ'' can be neglected at the minimum position, and M^0 is independent of H^0. This relation for $\omega_{\min}$, together with the resonance condition for maximum absorption, may be used to determine the g-factor without requiring a knowledge of M^0. In practice, however, the conditions are not easily satisfied, the minimum absorption often falling too near to zero field for M^0 to be taken independent of H^0.

The use of colloidal suspensions was proposed in an attempt to overcome some of the difficulties inherent in the use of plane samples and to provide an alternative verification of the g-factors and line widths since problems arise in connection with these measurements (see below). The colloids were obtained either by the reduction of chemical compounds or from a high-frequency arc run between electrodes of the metal to be examined. Experimentally, the particles seemed to be small compared with the skin depth and sufficiently spherical in shape so that the spectroscopic splitting factor could be derived directly from the relation $\hbar\omega_r = g\beta H^0$ (eqn. (6.10)) when anisotropy terms were neglected. The anisotropy terms can in this case be fairly rigorously treated (Bagguley 1955, Standley and Stevens 1956), and for narrow line widths the maximum absorption for a random assembly of spherical crystals occurs at a field $H^0 = \hbar\omega/g\beta + K_1/2M^0$.

For broader lines, the shift in the peak absorption is less than $K_1/2M^0$. Although this method has attractive features, the difficulties of obtaining sufficiently small particles (~ 100 Å) are considerable and it is not possible to obtain oriented single crystals. This latter is a serious disadvantage at low temperatures where the anisotropy energy increases and produces considerable shifts in the resonance peak (Bagguley and Harrick 1954).

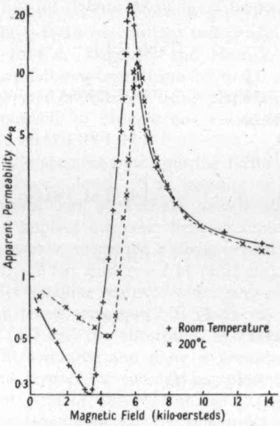


Figure 7.1. Ferromagnetic resonance absorption in a copper-nickel alloy containing 9·75 at % copper at 1·22 cm wavelength (*after* Standley and Reich 1955). $\mu_R = (|\mu| + \mu'')^{\frac{1}{2}}$, see text.

7.2. *Spectroscopic Splitting Factors*

The spectroscopic splitting factors obtained for various metals and alloys are collected together in table 7.1, together with values of g', the gyromechanical factor. The points of interest are (*a*) the validity of the relation $g-2 = 2-g'$ (see § 6), (*b*) the dependence if any of the g-factor upon ' electron concentration ' in alloys, saturation magnetization, temperature, or frequency.

Of the spectroscopic factors g given in table 7.1, that for nickel is the most satisfactorily established and agreed by workers in different laboratories. The determination of g from the resonance condition involves correction factors which may not be accurately known and so experiments in different laboratories using different experimental techniques and different frequencies are desirable before a value may be regarded as satisfactorily determined. Similar considerations apply to the determination of the gyromechanical factor g'. The experimental techniques are difficult and, moreover, Scott (1955 a, b) has reported that g' depends upon the induced intensity of magnetization in the sample. He obtains values of g' at low intensity of magnetization (Fe 1·87, Ni 1·80) which differ from the free spin value of 2·0 by larger amounts than those obtained at high intensity

(Fe 1·92, Ni 1·83). If the g-factor for nickel is accepted as well established, the relation $g-2=2-g'$ is not satisfied by the values of g' derived from the high intensity measurements (Brown, Meyer and Scott 1954). For other metals and alloys the situation is more obscure and the relation appears to be more or less satisfied in a number of cases. In general it may be said that qualitatively $g-2\simeq 2-g'$ but the quantitative verification is not yet complete. Fletcher (1954) has calculated g and g' for Ni in the approximation of tight binding. He concluded that $g-2$ should equal $2-g'$ on his model.

Table 7.1

Material (atomic %)		g (290°K)	g' (290°K)	Remarks
Fe		2·03–2·15 (a) 2·06 (b) 2·04–2·10 (e)	1·94, 1·94, 1·93 (n)	Low induced magnetization g' = 1·87
Ni		2·20 (f) 2·19 (d) 2·22 (b)	1·91, 1·93, 1·84 (n)	g independent of temperature 4°K–631°K. ? anomaly at 631°K. Low induced magnetization g' = 1·80 (p)
Co			1·86 (n)	
FeNi	Ni 40%–85%	2·10 (b)		? $g\sim 2·0$ near 40% Ni, g' = 1·98 near 30% Ni
	Ni 50%		1·91 (n)	
	Ni 36%–48%	2·08–2·14 (g)		? g frequency dependent
FeNi	Ni 15%–25%	2·06 (b)		
FeSi	Si 7·5%	2·0 ±0·2, 2·14 ±0·08 (h)		
	Si 0%–8·5%	2·06 (b)		
	Si 5%	2·09 (e)		
FeCr	Cr 14%, 19%	2·06 (b)		
CuNi	Cu 0%–28%	2·17–2·22 (d, j)		
MnNi	Mn 0%–14%	2·19 falling to 2·11 (d)		
AlNi	Al 0%–10%	2·17–2·21 (d)		
SbNi	Sb 0%–4%	2·17–2·19 (d)		
Supermalloy		2·12 (f)	1·91 (n)	
Permalloy		2·07–2·14 (k)	1·90 (n)	
Heusler alloy		2·01 (m)	2·00 (o)	
Gadolinium		1·94 (q)		g = 1·95 above Curie point

References: a, Griffiths 1946; b, Bagguley 1955; c, Bagguley and Harrick 1954; d, Standley and Reich 1955; e, Barlow and Standley 1956; f, Bloembergen 1950; g, Hoskins and Wiener 1954; h, Kip and Arnold 1949; j, Meyer 1952; k, Kip and Arnold 1948; m, Yager and Merritt 1949; n, Brown, Meyer and Scott 1954; o, Barnett 1944; p, Scott 1955; q, Kip 1953.

Measurements on alloy series have been reported by Meyer (1952) CuNi, Bagguley (1955) FeNi, FeSi, and Standley and Reich (1955) CuNi, MnNi, AlNi. It can be seen from table 7.1 that the g-factor appears to remain constant over wide ranges of alloy composition provided that the crystal structure does not change. Addition of the non-magnetic metals Cu, Al, Sb to Ni seems to leave the g-factor unchanged, whereas both Fe and Mn produce a sharp reduction in *g*-factor. Standley and Reich (1955) have suggested that their results on MnNi may be explained if the Mn and Ni atoms form separate sub-lattices for which the observed *g*-factor will be a weighted mean (see § 6). Bagguley (1955) in measurements on colloids of NiFe alloys found an anomaly near 40% Ni, where the *g*-factor fell to 2·01. Measurements on plane samples of alloys near this composition (Hoskins and Weiner 1954) have not confirmed the anomaly.

The spectroscopic splitting factor for nickel has been measured over a wide range of temperature (Standley 1949, Bloembergen 1950, Bagguley and Harrick 1954, Reich 1956). The *g*-factor is constant independent of temperature below the Curie point down to 4°K. Bagguley and Harrick (1954) measured nickel colloids and found a sharp fall in *g*-factor from 2·2 to 2·0 at the Curie point. This change has not been reported by workers using plane samples, where it does not appear to have been possible to observe any resonance absorption above the Curie point.

The dependence of the spectroscopic splitting factor upon frequency has not been investigated in sufficient detail. The majority of measurements on plane samples have been carried out at 1·25 cm wavelength, for which the peak absorption occurs at an applied magnetic field of about 5000 Oe. Barlow and Standley (1956) have, however, measured a single crystal of 2½% SiFe at 1·23 cm and 0·86 cm wavelength. They find $g = 2{\cdot}14$ (1·23 cm) and $g = 2{\cdot}10$ (0·86 cm). On the other hand, polycrystalline material which presumably will have a contribution from anisotropy terms gave $g = 2{\cdot}03$ (1·23 cm) and $g = 2{\cdot}06$ (0·86 cm). At longer wavelengths (~ 3 cm) the absorption line extends to zero field where the material may not be saturated and there is considerable uncertainty in the derived *g*-value. In this respect the colloidal samples offer some advantages but the changes in apparent *g*-factor observed so far (Bagguley and Harrick 1954, Bagguley 1955) may be accounted for by anisotropy corrections. Kittel and Mitchell (1956) have proposed that there is a fundamental frequency dependence of the *g*-factor (e.g. in the measurements of Hoskins and Weiner 1954) and they have suggested that it originates from coupling between the 3d and 4s electrons, cf. § 3.2. This produces a polarization of the conduction electrons leading to an extra internal field (eqn (3.6)) acting upon the 3d magnetization. It is suggested that such a field could account for discrepancies in the relation $g - 2 = 2 - g'$.

7.3. *Anisotropy Constants*

Measurements on single crystals enable the anisotropy constants of the material to be determined from the variation in the position of the peak absorption as a function of crystal orientation in the steady magnetic field. Theoretical discussions and formulae have been given by Kittel (1948) and Van Vleck (1950). Experimental investigations have been reported by Kip and Arnold (1949) for FeSi (7·5 at. % Si), Barlow and Standley (1956) for FeSi (5 at. % Si), and by Reich (1956) for Ni. The results verify the theoretical predictions and give values for the first order anisotropy constants which are in reasonable agreement with static methods. There is a discrepancy, however, in the value of K_2, the second order constant for nickel; here the microwave measurements give values opposite in sign from the static measurements. The results are listed in table 7.2.

Table 7.2

Material	Temp.	Microwave		Static†	
	(°K)	K_1	K_2	K_1	K_2
		($\times 10^{-3}$ erg cm^{-3})		($\times 10^{-3}$ erg cm^{-3})	
FeSi : (7·5 at. % Si)	290	300	—	320	—
(5 at. % Si)	290	370	(130 ± 20)	380	—
Ni	290	-46	$-(36 \pm 15)$	-50	$+50$
Ni	77	-540	$-(130 \pm 50)$	-600	$+70$
Ni	4·2	-830	$-(140 \pm 50)$	-750	$+100$

† Static values taken from Bozorth (1951).

7.4. *Line Shapes and Relaxation Terms*

The experimental absorption curves may be fitted to a theoretical equation of motion which uses either the Landau–Lifshitz or the Bloch relaxation terms (§ 6). When the measurements are analysed to give the imaginary part of the permeability μ_2, the line shape is of the Lorentz type ($P \propto 1/(1+x^2)$, cf. eqn (3.2)) and the line widths are of the order of hundreds of Oe corresponding to values of $T_2 \sim 10^{-9}$ sec.

Of the pure metals, only nickel has been investigated over an extensive temperature range. The line width appears to be independent of temperature below 500°K (Bloembergen 1950, Bagguley and Harrick 1954, Standley and Reich 1955, Reich 1956). Above 500°K measurements in plane specimens indicate a steady increase in line width up to the Curie point (631°K) whereas in colloids the line width does not appear to increase appreciably until the vicinity of the Curie point is reached. Standley and Reich (1955) have suggested that their measurements on alloys of Ni with non-magnetic metals (Cu, Al, Sb) may indicate that the line width is determined only by the saturation magnetization of the alloys. The temperature and constitution of the alloys are then significant only in so far as they modify the saturation magnetization. Bagguley (1955) endeavoured to correlate the anisotropy factor $K_1/2M^0$ with the line widths obtained for the series of colloidal FeNi alloys. The measurements suggest that anisotropy and other terms contribute more markedly to the line width in the case of the colloidal suspensions than in the case of plane samples and the colloid measurements yield somewhat greater line widths than are obtained from plane materials (e.g. Ni colloid $\Delta H_{\frac{1}{2}} \sim 480$ Oe, Ni plane $\Delta H_{\frac{1}{2}} \sim 150$ Oe at 290°K).

The physical origin of the relaxation mechanism is obscure and no satisfactory theory has been proposed. We shall not discuss the problem in detail since Kittel and Abrahams (1953) and Abrahams (1954) have reviewed the previous suggestions and concluded that the mechanisms proposed (e.g. spin–spin interactions, exchange effects in the eddy current skin depth, spin–lattice interactions, both direct and via conduction electrons) yield line widths considerably narrower than those observed in practice. The treatment is usually based on the spin-wave formulation of ferromagnetism and the difficulties arise because only long wave ($\mathbf{k} = 0$) spin-waves are excited by the resonance process. These can then only interact with the long wave lattice phonons and the relaxation time is much too long. Kittel and Abrahams (1953) endeavoured to overcome this difficulty by introducing the zero-point fluctuations of the spin-system and suggested that this mechanism could excite spin-waves of shorter wavelength, thus decreasing the relaxation time. It has not been possible to demonstrate that this process can in fact take place (Abrahams 1954). More recently Clogston *et al.* (1956) have suggested that relaxation times in a practical bounded sample may differ from those calculated for an infinite medium because of effects arising from the demagnetizing terms. They calculate that for a spheroid there will be a number of different spin-waves ($\mathbf{k} \neq 0$) which will have the same resonance condition as those with $\mathbf{k} = 0$. It is suggested that this degeneracy will give rise to a relaxation process depending linearly upon the excitation of $\mathbf{k} = 0$ spin-waves and linearly upon the temperature. The theory has not yet been published in sufficient detail for a critical discussion to be presented, but since the shape of the sample is included through the demagnetizing factors it is tempting to hope that this mechanism may account for the broader line widths measured in colloidal suspensions. Kittel and Mitchell (1956) have suggested that coupling between 3d–4s electrons may contribute to the line width as well as modify the apparent *g*-factor (§§ 7.2 and 3.2). This theory requires the conduction electron spin to be relaxed independently by a rapid mechanism (e.g. spin–orbit interaction) and a coupling constant between 3d–4s electrons which is only a few per cent of the free atom value.

§ 8. FERRITE MATERIALS

8.1. *General Discussion*

The ferrite materials which are of interest belong to the group of compounds having the general formula $MO.Fe_2O_3$, with M a divalent metal ion such as Mn, Co, Ni, Cu, Zn, Cd, Mg, Fe. They have the spinel structure, so called after the mineral Spinel $MgAl_2O_4$, but two different classes, ‘ normal ’ and ‘ inverted ’, are distinguished according to whether the divalent metal ion has tetrahedral or octahedral groups of surrounding oxygen atoms. The Mn, Fe, Co and Ni ferrites are inverted spinels, whereas the Zn, Cd ferrites are normal. Such materials have important practical applications at high frequencies because their large electrical resistance (of the order 10^2 to 10^7 ohm cm as compared with 10^{-5} ohm cm for iron) reduces eddy current losses. On the other hand, their saturation magnetization is rather low (of the order 500 gauss compared with 2000 gauss for iron) and is only about one-third of that which would be estimated assuming that all the metal atoms have their electron spins parallel. Néel (1948) suggested that this discrepancy arises because the spin system located on the octahedral sites is oriented antiparallel to that on the tetrahedral sites. It is, therefore, a type of antiferromagnetic material (see § 9) in which the moments of the two sub-lattices are not equal and so there remains a resultant magnetization. Néel proposed that the ferrites were particular examples of a more general type which he suggested should be called ferrimagnetic systems.

The properties of those ferrites which have useful technical applications have been reviewed by Fairweather, Roberts and Welch (1952) and by Smit and Wijn (1954). The ferrites are prepared by a sintering process which consists of firing the compressed materials at a temperature of about 1250°c in oxygen or in air. The porosity of the product depends upon the nature of the raw materials, thus manganous ferrites are usually rather dense whereas the porosity of nickel ferrite may amount to 25% (Smit and Wijn 1954). High densities may be obtained by firing the ferrite at a temperature above 1400°C, but in this case the equilibrium oxygen pressure is greater than one atmosphere and partial reduction may occur. It is therefore a problem of considerable magnitude to obtain reproducible sintered stoichiometric compounds with densities approaching that of the crystalline material. In principle, ferrites may be grown from the melt but this is not a satisfactory method on a commercial scale because of the high melting points (~1600°C) and high equilibrium oxygen pressures. Single crystals have been obtained on a small scale by this method, however, and also from solutions in molten borax and by flame fusion methods. Such crystals are frequently not very uniform in chemical composition and considerable care is necessary when selecting samples for magnetic measurements.

Because of the difficulties involved in growing satisfactory single crystals and the relative ease with which sintered samples may be obtained, the majority of microwave measurements have been carried out with these polycrystalline samples. Such measurements have provided useful information about particular applications, for example in the use of these materials as microwave attenuators (see § 8.3), or they may have been helpful in studying factors in the preparation of the ferrite which may influence the magnetic properties, for example the oxygen atmosphere, the firing time, or the firing temperature. They are, however, of limited value in the case of magnetic resonance measurements which are directed at understanding the fundamental physical processes. This is a particularly relevant consideration with regard to measurements which have been reported on materials with only a nominal chemical composition and for which the density, porosity and magnetic anisotropy are frequently unknown. Fortunately a number of significant measurements have recently been carried out with single crystal samples and we shall discuss only these experiments in this section of the report.

8.2. *Magnetic Resonance*

It appears that the most satisfactory experimental technique is that using a spherical sample (~0·03 cm diameter) which has dimensions small compared with the microwave wavelength in the material. The crystal is oriented by X-rays and mounted at the centre of a full wavelength T.E. mode cavity in order to reduce dipole image effects which may arise if the sample is situated near to a conducting wall (Yager *et al.* 1950, Bloembergen 1952). Otherwise the experimental techniques do not differ from the usual magnetic resonance arrangements and measurements have been reported at wavelengths between 10 cm and 1·25 cm and over the temperature range 4°K to 500°K (see table 8.1).

The spectroscopic splitting factor and the anisotropy terms may be obtained from equations (6.10) and (6.11). These results together with measurements of the line widths are summarized in table 8.1. Here we have not necessarily reproduced all the results reported over the whole temperature range but have selected typical values. For further details reference may be made to the original papers. The measured g-factor should be a weighted mean value for the system of coupled sub-lattices, as discussed in § 6. As yet, however, there are not sufficient experimental results from materials with accurately known chemical composition to make a critical assessment of the theory. In so far as any tentative conclusions

Table 8.1

Sample	Wavelength (cm)	T (°K)	K_1/M^0 (Oe)	K_2/M^0 (Oe)	M^0 (gauss)	g	$\Delta H_{\frac{1}{2}}$ (111) (Oe)	Remarks
$Mn_{0·98}Fe_{1·86}O_4$	5·1, 3·2, 1·25	300	−79 ± 3	6 ± 10	351	2·004	33	g independent of frequency 290°K. Angular variation of $\Delta H_{\frac{1}{2}}$, 290°K (111) 33 Oe, (110) 25 Oe, (100) 23 Oe (Dillon *et al.* 1955)
		77	−412 ± 5	−5 ± 10	493	2·02	80	
$\rho \simeq 10^4$ ohm cm		4·2	−433 ± 15	−200		(2·06)	90	
$MnZnFe_2O_4$	10, 3·3, 5·3, 1·25	300	−70			2·00	~50	g independent of frequency and temperature 300°K–4·2°K (Tannenwald 1955 b)
$\rho \simeq 10^4$ ohm cm		4·2	−750			2·00	~200	
$MnFeFe_2O_4$	10, 3·3, 5·3, 1·25	300	−70			2·00	50–250	g independent of frequency and temperature 300°K–4·2°K. Line width frequency dependent 300°K; independent of frequency 4·2°K. (Tannenwald 1955 b)
$\rho \simeq 10$ ohm cm		4·2	−250			2·00	~100	
$Mn_{0·45}Zn_{0·55}Fe_2O_4$	1·25	290	−21 ± 1		181 ± 20	2·0	~35	Galt *et al.* 1951
$(NiO)_{0·95}(FeO)_{0·05}Fe_2O_3$	1·25	290	−254	−70 ± 50	270	2·2	50	Measurements 290°K–4·2°K. Angular variation $\Delta H_{\frac{1}{2}}$ is small (Yager *et al.* 1955)
		85	−370	−100 ± 50	290	2·2	55	
$\rho \sim 1$ ohm cm		4·2	−392	−120 ± 50	300	2·2	30	
$(NiO)_{0·75}(FeO)_{0·25}Fe_2O_3$	1·25	290	−152	16	322	2·13	140	Measurements 290°K–4·2°K. $\Delta H_{\frac{1}{2}}$ shows maximum at about 160°K. Some angular variation in $\Delta H_{\frac{1}{2}}$ (Yager *et al.* 1955)
		85	−259	24	340	2·19	120	
$\rho \sim 1$ ohm cm		4·2	−421	−15	342	2·16	40	
$NiOFe_2O_3$	1·25	290	−236		265	2·19	~35	Angular variation $\Delta H_{\frac{1}{2}}$ less than variations between different crystal samples (Yager *et al.* 1950)
$CoOFe_2O_3$	1·25	553	0			2·27		Double resonance observed (Tannenwald 1955 a)
		363	2200			2·7		
$CuOFe_2O_3$	1·25	723	0		30	2·05		g increases as temperature decreases below 300°K, g constant above 300°K (Okamura and Kojima 1952)
		300	220		280	2·04		
		78	−650		315	2·20		
Fe_3O_4	3·3, 1·25	295	−230		472	2·12		Crystal disc, g decreases gradually with decreasing temperature $\lvert K_1 \rvert$ zero at 130°K (Bickford 1950)
		173	−200		491	2·08		
		150	+84		499	2·03		

may be drawn, the present results do not appear to conflict with the theory and the spectroscopic state of the magnetic atoms Mn^{2+}, Fe^{3+}, Ni^{2+} are similar to those found in the usual octahedral complexes for iron group salts (Bowers and Owen 1955). There appears to be no conclusive evidence for any variation of the g-factor with either frequency or temperature. The observed line widths, which are more sensitive to the microscopic composition of the crystal than are the g-factors, may vary more widely among specimens having apparently the same chemical composition than do the angular variations within a given crystal. The crystallographic direction for which the line width is a minimum is probably the (100) direction (Yager *et al.* 1955, $NiOFe_2O_3$; Dillon *et al.* 1955, $MnOFe_2O_3$) but some discrepancies have been observed (Tannenwald 1955 b). The line width is apparently temperature dependent and one case of frequency dependence has been reported (Tannenwald 1955 b).

8.3. *Relaxation Effects*

Relaxation effects in several ferrites have been investigated by Bloembergen and his co-workers (Damon 1953, Bloembergen and Wang 1954). The experiments were planned on the assumption that the changes in the magnetization $(M_z - M^0)$ and the imaginary part of the permeability μ'' could be derived from the Bloch type of equations (§ 6). As the microwave power is increased to saturate the absorption line, the relations between the amplitude of the microwave magnetic field h^0, the relaxation times T_1, T_2 and μ'' or M_z at resonance are

$$\mu''/\mu_0'' = M_z/M_0 = \{1 + \tfrac{1}{4}(\gamma h^0)^2 T_1 T_2\}^{-1}. \qquad \ldots\ldots(8.1)$$

μ_0'' is the value of μ'' at resonance with no saturation, i.e. as $h^0 \rightarrow 0$. The ratios μ''/μ_0'' and M_z/M^0 should therefore lie on the same curve. This was verified to be true for the paramagnetic materials $MnSO_4 4H_2O$ and α–α–diphenyl–β–picrylhydrazyl. However, the experimental values for μ''/μ_0'', M_z/M^0 did not lie on the same curve in the case of ferrimagnetic materials at temperatures well below the Curie point. For example, in $NiOFe_2O_3$ at 300°K, it was found that the amplitude of the microwave magnetic field required to reduce μ''/μ_0'' by one half was 100 times less than that required to produce the corresponding reduction in M_z/M^0. These anomalous effects gradually decreased as the temperature was increased to the Curie point. There does not appear to be a satisfactory explanation of these phenomena but it is possible to derive the relaxation time τ_s for energy exchange within the spin system by a general thermodynamic argument from $M_z - M^0$ without having recourse to the particular model of eqn (8.1) (Anderson 1952, Bloembergen and Wang 1954). At high temperatures τ_s obtained by this method was found to correspond to the inverse absorption line width, whereas at low temperatures τ_s increased roughly in inverse proportion to the temperature without any corresponding change in line width. The order of magnitude of τ_s for the ferrites $NiOFe_2O_3$, MnO (14·9%) ZnO (15·6%) FeO (1·5%) Fe_2O_3 (67·9%), NiO (35%) ZnO (15%) Fe_2O_3 (50%), is 2×10^{-9} sec at 300°K and 7×10^{-9} sec at 77°K. In NiO (11·4%) ZnO (38·6%) Fe_2O_3 (50%) no saturation effects were observed at 77°K, whilst at 300°K τ_s was 2×10^{-9} sec. The measurements were extended to Supermalloy for which $\tau_s \sim 10^{-9}$ sec and nickel metal for which no saturation was observed.

8.4. *Faraday Rotation*

Systems which employ the rotation of the plane of polarization that occurs when a plane wave is propagated through a magnetic medium, e.g. a ferrite, are important as one-way transmission elements in waveguide circuits. Detailed discussions of these non-reciprocal devices using longitudinal and transverse magnetization with different geometric forms of material have been given by Suhl and Walker (1954 a, b, c). The practical applications in the form of directional phase shifters, resonance and resistance isolators, and circulators of various types have been described by Hogan (1953), Rowen (1953) and Fox, Miller and Weiss (1955). The presence of waveguide walls has a considerable effect on the quantitative relations between rotation of the plane of polarization or wave velocity, and the magnetization, but the principle of the phenomena is not different from that which occurs in an unbounded medium (Polder 1949). The theoretical discussion will therefore be limited to this case.

We consider an infinite non-conducting medium, without anisotropy terms and neglecting relaxation effects. Then $H_1 = H_2 = H^0$, all the N's are zero in eqn (6.6) and $\mu_1 = \mu_2$. When a circularly polarized wave is propagated through such a medium in the direction of H^0 (positive z)

$$\text{for } h_x = +jh_y, \quad \mu^+ = 1+\left(\frac{4\pi m}{h}\right) = 1+\frac{4\pi\gamma M^0}{\gamma H^0+\omega}, \qquad \ldots\ldots(8.2)$$

$$\text{for } h_x = -jh_y, \quad \mu^- = 1+\frac{4\pi\gamma M^0}{\gamma H^0-\omega}. \qquad \ldots\ldots(8.3)$$

A plane polarized wave can be resolved into two oppositely rotating circularly polarized waves and these will have different velocities v^+, v^- corresponding to the permeabilities μ^+, μ^- in equations (8.2) and (8.3). The angular rotation of the plane of polarization θ, in a distance l, is given by

$$\frac{\theta}{l} = \frac{\omega}{2}\left(\frac{1}{v^-}-\frac{1}{v^+}\right) \qquad \ldots\ldots(8.4)$$

so that taking the dielectric constant ϵ to be real,

$$\frac{\theta}{l} = \frac{\omega\sqrt{\epsilon}}{2c}\left[\left\{1+\frac{4\pi\gamma M^0}{\gamma H^0-\omega}\right\}^{1/2} - \left\{1+\frac{4\pi\gamma M^0}{\gamma H^0+\omega}\right\}^{1/2}\right]. \qquad \ldots\ldots(8.5)$$

For $\omega \gg \gamma H^0$ and $\omega \gg 4\pi\gamma M^0$ this reduces to

$$\frac{\theta}{l} = \frac{2\pi\gamma M^0\sqrt{\epsilon}}{c}. \qquad \ldots\ldots(8.6)$$

In small magnetizing fields therefore the rotation of the plane of polarization is proportional to the magnetization of the medium and independent of the frequency. For typical values (Hogan 1953) of $\epsilon \sim 10$, and $4\pi M^0 \sim 3000$ gauss, $\theta/l \sim 160^\circ$/cm. In a waveguide the system is no longer independent of frequency and moreover the dependence of M^0 on temperature may introduce practical difficulties. Fox, Miller and Weiss (1955) quote a ferrite with a Curie temperature of 150°c as giving a temperature variation of 0·4% change in rotation per centigrade degree at room temperature, but this effect is much reduced by using materials with a higher Curie point.

When the magnetization is perpendicular to the direction of propagation a reciprocal element is obtained when the ferrite is situated symmetrically within the waveguide. The amount of phase shift can be controlled by an applied magnetic field. For a wave polarized along the steady field H^0 the effective permeability is just the high frequency permeability, whilst for a component polarized perpendicular to H^0 the permeability is given by (Polder 1949, Weiss and Fox 1952)

$$\mu_\perp = \frac{\mu^2 - k^2}{\mu} = \left\{1 + \frac{16\pi^2\gamma^2(M^0)^2 + 4\pi\gamma^2 M^0 H^0}{4\pi\gamma^2 M^0 H^0 + \gamma^2(H^0)^2 - \omega^2}\right\} = \frac{\gamma^2(H^0 + 4\pi M^0)^2 - \omega^2}{\gamma^2 H^0(H^0 + 4\pi M^0) - \omega^2}. \qquad \ldots\ldots(8.7)$$

This expression is independent of the direction of propagation and so the transverse magnetic field produces reciprocal birefringence. The effect is small except at frequencies for which $\omega \sim \gamma\{H^0(H^0 + 4\pi M^0)\}^{1/2}$. If $\omega \gg \gamma H^0$ and $\omega \gg 4\pi\gamma M^0$ the relative phase retardation of the wave with its magnetic vector parallel to H^0 compared with the wave whose magnetic vector is perpendicular to H^0 is given by

$$\frac{\theta'}{l} = \frac{\sqrt{\epsilon}}{2}\left\{\frac{\gamma^2 4\pi M^0(H^0 + 4\pi M^0)}{\omega}\right\}. \qquad \ldots\ldots(8.8)$$

Non-reciprocal devices are constructed by placing the ferrite material at an asymmetrical position within the waveguide. These have useful applications as phase shifters and attenuators. It can be seen from equation (8.7) that $\mu_\perp$ becomes zero and changes sign in low magnetic fields where $\gamma(H^0 + 4\pi M^0) = \omega$ so that large changes in the wave velocity occur near this field strength. Phase shifts of about 170° between the forward and reverse directions of propagation have been observed with fields of about 500 oersteds (Rowen 1953). This low magnetic field strength is advantageous for phase shifting devices because the insertion loss may be much less than that characteristic of the high field region where resonant power absorption occurs. It is therefore possible to obtain a variable phase shifter operating at low magnetic fields with reasonably constant attenuation. At the higher field strengths where $\gamma[H^0(H^0 + 4\pi M^0)]^{1/2} \sim \omega$ the asymmetrically loaded waveguides may be used as non-reciprocal attenuators. As such they may have about 0·5 dB forward loss and 20 dB reverse loss as compared with 0·1 dB and 30 dB for longitudinal Faraday rotation devices (Rowen 1953), but the transverse field devices are probably the more simple in construction. As differential phase shifters transverse field devices have about the same insertion loss as those employing Faraday rotation.

8.5. *The Initial Permeability*

Despite their high resistivity ferrites often exhibit considerable losses which limit their usefulness at high frequencies. These losses are thought to arise from the fundamental processes of magnetization. In the case of a ferromagnetic metal the initial magnetization is determined by the displacement of the Bloch walls which bound the individual domains and the contribution from the rotation of the magnetization within each domain is thought to be small because of the relatively high crystal anisotropy energy. On the other hand, in the sintered polycrystalline ferrites which have considerable porosity, the domain walls stretching across the pores may not be so free to move as in a metal and, moreover, the anisotropy energy may be much smaller (for example $MnFe_2O_4$ has $K_1 \sim 10^3$ erg cm^{-3} compared with $K_1 \sim 10^5$ erg cm^{-3} for iron). Thus, if the initial permeability is written in the form $\mu = 1 + 4\pi(\chi_{\text{disp}} + \chi_{\text{rot}})$ both the susceptibility terms will make significant contributions and their frequency dependence will determine the usefulness of the ferrite. In particular, Bloch wall or rotational resonance will give rise to associated losses and also set an upper limit to the range in which the material has an appreciable high-frequency permeability.

8.5.1. *Wall displacements.*

Within a Bloch wall the elementary spins are not completely parallel but precess at some angle to the effective field within the wall. Becker (1951) and Doring (1948) have pointed out that the wall will appear to have inertial mass which can be derived by equating $\frac{1}{2}m_w v^2$ with the excess energy of a moving wall over one that is stationary. There will also be restoring forces binding the wall to its equilibrium position and dissipative forces which appear as damping terms when the Bloch wall is in motion. Under the action of an oscillatory field $h(t)$, therefore, the equation describing the Bloch wall movements will be of the form

$$m_w \ddot{x} + \beta \dot{x} + \alpha x = 2M^0 h(t). \qquad \ldots\ldots(8.9)$$

This may lead to a resonance or to a dispersion in the permeability with associated losses according to the values of the coefficients, m_w, β, α. We estimate their order of magnitude in the discussion below.

The rotational motion of the local magnetization which accompanies the uniform motion of a Bloch wall (with velocity v in the z direction) may be supposed to be caused by an effective field $H_e(z)$ such that the precessional velocity $d\phi(z)/dt$ is equal to the rotational velocity required by the motion of the wall, i.e. $\gamma H_e = d\phi/dt = -v\, d\phi/dz$. In a moving wall the additional energy may be written in terms of the field H_e

$$\Delta W \equiv \tfrac{1}{2}m_w v^2 = \int_{-\infty}^{+\infty}\left(\frac{H_e^2}{8\pi}\right)dz = \frac{1}{8\pi}\frac{v^2}{\gamma^2}\int_{-\infty}^{+\infty}\left(\frac{d\phi}{dz}\right)^2 dz. \qquad \ldots\ldots(8.10)$$

The integral $\int_{-\infty}^{+\infty}\left(\frac{d\phi}{dz}\right)^2 dz$ occurs in the elementary theory of the Bloch wall and is equal to $\sqrt{(K_1/A)}$ (Kittel 1949), where K_1 is the first order anisotropy constant, $A = 2JS^2/a$ with a the lattice constant, S the spin of each carrier and J the exchange integral. The effective inertial mass is therefore

$$m_w = \frac{1}{4\pi\gamma^2}\sqrt{(K_1/A)} \sim 3 \times 10^{-16}\sqrt{(K_1/A)}.$$

In terms of the wall thickness parameter $\delta \sim 4\sqrt{(A/K_1)}$, m_w is equal to $(\pi\gamma^2\delta)^{-1}$. For iron $\sqrt{(K_1/A)} \sim 10^6$ cm^{-1} so that $m_w \sim 10^{-10}$ g cm^{-2}. In ferrites the values for m_w range from 10^{-10} to 10^{-11} g cm^{-2} (Galt *et al.* 1953, Smit and Wijn 1954).

The relaxation process which maintains the spins in thermal equilibrium with the lattice introduces a dissipative force during the motion of the Bloch wall.

The force may be written as βv, and the corresponding energy dissipation per unit area of the wall $W_{\mathrm{D}} = \beta v^2$. The energy dissipated per unit volume during the magnetization process is $\mathbf{H}^0 . d\mathbf{M}/dt$ so that

$$W_{\mathrm{D}} = \int \mathbf{H}^0 . \frac{d\mathbf{M}}{dt} dz \sim \int \lambda H_{\mathrm{e}}^2 dz \qquad \ldots\ldots(8.11)$$

when the relaxation term is of the Landau–Lifshitz form (see § 6). The coefficient β is then obtained from equation (8.10), $\beta = 4\pi\lambda m_{\mathrm{w}}$.

The Bloch wall is also considered to be bound to its equilibrium position by an elastic force, which may be written as $-\alpha x$ per unit area of the wall for a given displacement x. If the magnetizing field is parallel to the magnetization in one domain and we consider a 180° wall, the force acting on the wall is $2M^0H^0$ so that $x = 2M^0H^0/\alpha$. The change in magnetization produced by a wall movement is $2M^0x/l$ where l is the average distance between the domain walls, so that the initial susceptibility is related to the elastic constant α by the equation

$$\chi_{\mathrm{disp}}{}^0 \sim \frac{2M^0 x}{lH^0} = \frac{4(M^0)^2}{\alpha l}. \qquad \ldots\ldots(8.12)$$

When β is small ($\beta^2 \ll 4\alpha m_{\mathrm{w}}$) a distinct resonance in the permeability should occur at a frequency $\omega = \sqrt{(\alpha/m_{\mathrm{w}})}$. For iron with

$$M^0 \sim 1000 \text{ gauss}, \ \chi_{\mathrm{disp}}{}^0 \sim 10, \ l \sim 10^{-3} \text{ cm},$$

ω is about 10^9 c/s (Johnson and Rado 1949). For ferrites taking

$$\alpha \sim 10^5, \ m_{\mathrm{w}} \sim 10^{-11} \text{ g cm}^{-2},$$

ω is about 10^8 c/s. Rado (1953) has reviewed the frequency dependence of the initial susceptibility in ferrites. He and his collaborators have reported two resonance peaks in MnMg ferrite and they ascribe the peak at the lower frequency (~ 50 Mc/s) to Bloch wall movements because this peak can be eliminated by using particles which are sufficiently small to behave like single domains. Wijn (1953) has also reported a wall resonance in $\mathrm{BaO.6Fe_2O_3}$ (Ferroxdure) at a frequency of about 200 Mc/s.

When λ is large, $\beta \gg m_{\mathrm{w}}$, the wall displacement for an exciting field of the form $\exp(j\omega t)$ is obtained from equation (8.9), $x = 2M^0h/(\alpha + j\omega\beta)$. The susceptibility is therefore obtained from equation (8.12)

$$\chi \sim \frac{\alpha\chi_{\mathrm{disp}}{}^0}{\alpha + j\omega\beta}. \qquad \ldots\ldots(8.13)$$

The permeability will show a frequency variation of the relaxation type with a characteristic time $\tau_{\mathrm{D}} \sim \beta/\alpha$. Galt *et al.* (1953) have given values of β and α for $\mathrm{Fe_3O_4}$ of 0·5 and 10^4; and for $\mathrm{NiOFe_2O_3}$ of 0·02 and 10^6. The characteristic time τ_{D} is therefore of the order 10^{-4} sec in these cases.

8.5.2. *Rotational effects.*

The rotational susceptibility of ferromagnetic materials is determined largely by the crystalline anisotropy terms. From equation (6.7) it can be seen that there will be an anomaly in the initial permeability and an increase in the losses at a frequency $\hbar\omega_r \sim g\beta N^A M^0$ when demagnetizing effects are neglected. In the case of a uniaxial crystal, with the oscillatory field perpendicular to the crystal axis, $N_x{}^A = N_y{}^A \sim 2K_1/(M^0)^2$ (Kittel 1948) giving $\hbar\omega_r \sim g\beta(2K_1/M^0)$. For $NiOFe_2O_3$, Yager *et al.* (1950) give $K_1/M^0 \sim 240$ Oe so that $\omega_r \sim 10^9$ c/s. Snoek (1948) and Rado (1953) have ascribed the losses which occur in ferrites at frequencies above 100 Mc/s to resonance phenomena of this type. Typical values of ω_r obtained experimentally are Ferroxcube (MnZn, NiZn ferrites) 10–500 Mc/s, Ferramic (MnMg ferrite) 2000 Mc/s.

Kittel (1951) has given the frequency variation of the initial rotational susceptibility as derived from a simplified form of equation (6.4) with a Landau–Lifshitz relaxation term. He obtains

$$\frac{\chi_{\mathrm{rot}}(\omega)}{\chi_{\mathrm{rot}}{}^0} = \frac{\omega_0{}^2 + (\lambda/\chi')(j\omega + \lambda/\chi')}{\omega_0{}^2 + (j\omega + \lambda/\chi')^2}$$

in which $\omega_0{}^2 = \gamma^2(2K/M^0)^2 + (\lambda/\chi^1)^2$ and χ^1 is an average rotational susceptibility approximately equal to $2\chi_{\mathrm{rot}}{}^0[\sim 2(M^0)^2/3K_1]$. Experimentally, the losses extend over a much greater frequency range than would be predicted by the Landau–Lifshitz formula and there may be other contributions to the resonance condition which influence the experimental results (Abrahams 1954). The observed frequency variation seems to be due to a true resonance phenomenon since the real part of the initial permeability μ' shows a maximum and also $\mu' - 1$ becomes negative at high frequencies (Wijn *et al.* 1953). Kittel (1951), however, has pointed out that for $\lambda \gg \gamma M^0$ the frequency variation will be of the relaxation type

$$\frac{\chi_{\mathrm{rot}}(\omega)}{\chi_{\mathrm{rot}}{}^0} \sim \frac{1}{1 + j\omega(M^0)^2/3K_1\lambda}.$$

The characteristic time is $\tau_r \sim (M^0)^2/3K_1\lambda$, which for $NiOFe_2O_3$ with $M^0 \sim 240$ gauss, $K_1/M^0 \sim 240$ Oe (Galt *et al.* 1953, Yager *et al.* 1950) gives $\tau_r \sim 10^{-7}$ sec.

Polder and Smit (1953) have shown that if the radio-frequency field has a component parallel to the Bloch wall, poles will be created which, together with demagnetizing fields, can modify the effective anisotropy field H^A acting on the spins. The resonance frequency is then given by $\hbar\omega_r = g\beta\sqrt{\{(H^A + 4\pi M^0)(H^A + N_y M^0)\}}$ for the radio-frequency field along the y axis. If $N_y = 4\pi$, then

$$\hbar\omega_r = g\beta(H^A + 4\pi M^0)$$

which will be a much higher frequency than would be predicted without these poles. Above this frequency losses in ferrites should be small and there is some confirmation of this from the experiments of Beljers, Van der Lindt and Went (1951).

8.5.3. *Maximum usable frequency*.

It seems likely that Bloch wall displacements should cease to contribute to the high-frequency permeability before the rotational terms become negligible. The maximum usable frequency of a ferrite is then of the order of magnitude of the rotational resonance frequency $\hbar\omega_r \sim g\beta(2K_1/M^0) \sim 2g\beta M^0/3\chi_{\mathrm{rot}}{}^0$. In terms of the permeability, $\mu_0 = 1 + 4\pi\chi_{\mathrm{rot}}{}^0$ we obtain $\nu_{\max}(\mu^0 - 1) = \frac{4}{3}(g\beta/\hbar)M^0$ (Snoek 1948). It is temperature dependent since μ_0 and M^0 are functions of temperature.

For ferrites with high Curie points the losses do not necessarily diminish with increasing temperature since the anisotropy constant does not decrease over the whole range.

§ 9. ANTIFERROMAGNETIC RESONANCE

The general properties of antiferromagnetics have been reviewed by Van Vleck (1951), Lidiard (1954) and Nagamiya, Yosida and Kubo (1955), and we will give only a brief outline here. In the simplest case an antiferromagnetic consists of two sub-lattices of paramagnetic ions, A and B, which at low temperatures are magnetized in opposite directions by exchange interactions. The system can be represented diagrammatically by

$$\begin{array}{cccccccc} \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \mathrm{A} & \mathrm{B} & \mathrm{A} & \mathrm{B} & \mathrm{A} & \mathrm{B} & \mathrm{A} & \mathrm{B} \end{array}$$

On a molecular field approximation, each sub-lattice may be considered to be interacting with two molecular magnetic fields, H_E and H_a, where H_E is usually very much greater than H_a. H_E represents the main (isotropic) antiferromagnetic A–B exchange interaction, and on the molecular field model the field acting on sub-lattice A is $\mathbf{H}_E = -\lambda\mathbf{M}_B$, and on B is $\mathbf{H}_E = -\lambda\mathbf{M}_A$, where λ is a molecular field constant, and M_A, M_B are the spontaneous magnetizations of sub-lattices A and B. The order of magnitude of H_E at zero temperature is

$$H_E^0 \sim kT_N/\beta \qquad \ldots\ldots(9.1)$$

where T_N is the antiferromagnetic transition (or Néel) temperature, and β is the Bohr magneton which is the order of magnitude of the magnetic moment of each ion. For example, in $\mathrm{MnF_2}$, where $T_N \simeq 68°\mathrm{K}$, $H_E \sim 500$ kOe at 0°K. At intermediate temperatures, H_E decreases to zero as T increases to T_N. This follows the sub-lattice spontaneous magnetization which varies with temperature according to a modified Brillouin function (cf. ferromagnetism). The other molecular field H_a represents the anisotropy energy. If there is axial symmetry, for instance, the A and B spins like to be aligned along a particular direction in the crystal for a number of reasons which could include dipole–dipole coupling and anisotropic exchange effects arising from asymmetrical crystal fields (for a review, see Nagamiya, Yosida and Kubo 1955). There is no simple general expression for the value of H_a, but in typical cases it is of order 10^3 to 10^4 Oe at $T = 0$, and its temperature dependence is found in most cases to be similar to that of H_E.

The most important result in the theory (Kittel 1951, Nagamiya 1951) of the electron spin resonance from such a system is that, so far as its resonance properties are concerned, each sub-lattice sees an internal field†

$$H' \simeq (2H_E H_a)^{\frac{1}{2}} \qquad \ldots\ldots(9.2)$$

where it is assumed, as is usually the case, that $H_E \gg H_a$. As the temperature is lowered through T_N, there is then a magnetic field shift in the position of the paramagnetic resonance line of order H', and with zero applied field there is a splitting of the energy levels $\Delta E \sim g\beta H'$, where the spectroscopic splitting factor g is usually close to 2. This shift or splitting increases as the temperature is further lowered roughly according to a Brillouin function, reaching a saturation value at $T = 0$. For the example of MnF_2, $(2H_EH_a)^{\frac{1}{2}} \simeq 10^5$ Oe at $T = 0°\text{K}$, which corresponds to a zero field splitting of approximately $10\ \text{cm}^{-1}$. Thus, for this fairly typical case, resonance absorption at low temperatures can only be observed by using very high magnetic fields or very short wavelengths. Furthermore, since the position of the resonance line depends in an important way on the anisotropy field H_a, most antiferromagnetic resonance spectra show large anisotropy, and for successful experiments single crystals are normally essential. Finally, because of the temperature dependence of $(2H_EH_a)^{\frac{1}{2}}$, the positions of the absorption lines vary with temperature, and for any complete investigation good temperature control is necessary. Apart from the emphasis on short wavelengths, single crystals and temperature control, the experimental techniques for measuring antiferromagnetic resonance are the same as for normal paramagnetic resonance.

---

† We use the notation introduced by Kittel (1951). The important quantity $2H_E H_a$ has also been written in the following forms in other publications: (i) $2H_E H_a = c$; (ii) $2H_E H_a = 2\lambda K$ where K = anisotropy energy per unit volume $= H_a(M_A + M_B)/2 = H_a\overline{M}$, λ = molecular field constant $= H_E/\overline{M}$, and $\overline{M}$ = spontaneous magnetic moment of each sub-lattice per unit volume; (iii) $2H_E H_a = \alpha H_c^2$ or αH_c^{*2}, as suggested by eqn (9.10) below (see, for example, Nagamiya, Yosida and Kubo 1955, and Gerritsen 1955).

9.1. *Summary of Experiments*

For the reasons just given, early attempts (Trounson, Bleil, Wangsness and Maxwell 1950) to observe antiferromagnetic resonance in powdered Cr_2O_3 ($T_N \simeq 307°\text{K}$) using 3 cm wavelength, only showed that the resonance absorption line broadened out and disappeared as T was lowered through T_N. The same effect was found in similar experiments by other authors, for example by Hutchison in MnF_2, and by Maxwell and McGuire (1953) in Cr_2O_3, MnF_2, MnTe, MnS, MnO, and other oxides.

Successful antiferromagnetic resonance experiments up to the present time are fairly few, and are listed below. All of the results are for single crystals unless stated otherwise.

(i) $CuCl_22H_2O$, $T_N = 4{\cdot}3°\text{K}$. The low transition temperature implies that the internal fields are fairly small, and antiferromagnetic resonance can be observed using 3 cm wavelength. The symmetry is orthorhombic, and with the applied magnetic field along the crystal *a*-axis, which is the easy or preferred axis of magnetization, the Leiden group (see below) find from their resonance measurements down to about 1°K that $(2H_EH_a)^{\frac{1}{2}}$ increases to a saturation value of about 7 kOe as the temperature is lowered. They also find that the temperature dependence can be approximately described by a Brillouin function. The first measurements of antiferromagnetic resonance in this salt were reported by Poulis, van den Handel, Ubbink, Poulis and Gorter (1951), and further work at 3 cm wavelength by Ubbink (1953), Ubbink, Poulis, Gerritsen and Gorter (1953), and at 8 mm wavelength by Gerritsen, Okkes, Bölger and Gorter (1955), and Gerritsen (1955). A good review of both theory and experiment is given by Nagamiya, Yosida and Kubo (1955). The line widths are discussed by Gerritsen, Garber and Drewes (1956), and the intensities by Gerritsen and Garber (1956). This work on $CuCl_22H_2O$ provides the most complete set of antiferromagnetic resonance data at present available. Some of the results are shown in figure 9.2, and are further discussed below.

(ii) $CuBr_2 2H_2O$, T_N between 5 and 6·5°K. The structure is similar to that of $CuCl_2 2H_2O$, and preliminary measurements at 3 cm wavelength by Date (1956) suggest that the antiferromagnetic resonance spectrum is also similar but with a slightly stronger molecular field.

(iii) Cr_2O_3, $T_N = 307{\cdot}5$°K. When $T \ll T_N$, the saturation value of $(2H_E H_a)^{\frac{1}{2}}$ is expected to be very large, and wavelengths of order 1 mm might be required to observe resonance. However, as T is raised, $(2H_E H_a)^{\frac{1}{2}}$ becomes smaller and rapidly decreases to zero at the transition temperature. Antiferromagnetic resonance has been measured in this region (306·5 to 307·5°K) by McGuire, Scott and Grannis (1956) using 1·2 cm wavelength. They find that in this small temperature range $(2H_E H_a)^{\frac{1}{2}}$ varies approximately as $(\overline{M})^{1\cdot 75}$ where $\overline{M}$ is the sub-lattice spontaneous magnetization and is expected to show a temperature dependence like a Brillouin function for $S = 3/2$. Their results are shown in figure 9.3, and are further discussed below. A preliminary account of measurements on this salt over a wider frequency range (0·4 cm to 1·2 cm) is given by Dayhoff (1956).

(iv) MnF_2, $T_N = 67{\cdot}7$°K. Measurements have been made by Johnson and Nethercot (1956) over the temperature range 42 to 64°K, and with wavelength range 1·4 to 3·1 mm. With the d.c. magnetic field parallel to the tetragonal axis of symmetry of the crystal, they find that $(2H_E H_a)^{\frac{1}{2}}$ increases as the temperature is decreased approximately like a Brillouin function for $S = 5/2$, and that at $T = 0$, $(2H_E H_a)^{\frac{1}{2}} \simeq 90$ kOe. The results are in reasonable agreement with calculations of Keffer (1952), who had predicted that at $T = 0$, $H_E = 540$ kOe, $H_a = 8{\cdot}8$ kOe giving $(2H_E H_a)^{\frac{1}{2}} = 97$ kOe. The results are further discussed below.

(v) CuMn and AgMn alloys. T_N depends on Mn concentration, for example, $T_N \simeq 37$°K with 5·6 at. % Mn in Cu. Measurements down to 2°K with 1·2 and 3·3 cm wavelength have been made by Owen, Browne, Knight and Kittel (1956), and Owen, Browne, Arp and Kip (1957). With 5·6% Mn in Cu, for example, the results suggest $(2H_E H_a)^{\frac{1}{2}}$ increases to a saturation value of about 2·7 kOe as T is reduced to zero. H_a appears to be very small, less than 50 Oe, and resonance can be seen in polycrystals as well as in single crystals. As pointed out by the above authors, there is some doubt as to whether the usual antiferromagnetic resonance theory is applicable, and it is possible that the observed shift in the resonance line is in part due to the magnetization of the conduction electrons.

9.2. *Theory and Comparison with Experiment*

The theory of antiferromagnetic resonance was initiated by Kittel (1951) and Nagamiya (1951). Contributions to the more detailed theory have been made by a number of authors, including the Leiden group, with reference to their measurements on copper chloride (see references above and also Gorter and Hantjes 1952), but here we cite only two fairly general articles, one by Keffer and Kittel (1952) and the other by Nagamiya, Yosida and Kubo (1955) in their review on antiferromagnetism. The latter authors emphasize the case of orthorhombic symmetry and the application of the theory to $CuCl_2 2H_2O$. For the present purpose we consider the simpler case of uniaxial symmetry, with the z-axis as the easy or preferred axis of sub-lattice magnetization. At the risk of some oversimplification, we illustrate some of the resonance conditions with experimental results on $CuCl_2 2H_2O$, though this salt is really orthorhombic. We will quote formulae freely from the two papers just mentioned. For simplicity, it is assumed throughout that the intrinsic g-value of the paramagnetic ions is isotropic.

9.2.1. *H perpendicular to easy axis z.*

The system is shown in figure 9.1. The anisotropy fields H_a are directed along z, $-z$, so that with no applied field the magnetic moments of the sub-lattices M_A, M_B are directed along these ' easy ' axes to give the lowest anisotropy energy. If an external field H is now applied along x, M_A is acted on by H along x, by H_a along z, and by an exchange field $\mathbf{H}_E = -\lambda\mathbf{M}_B$ where λ is a positive molecular field constant. Similarly, M_B sees H along x, H_a along $-z$, and an exchange field $\mathbf{H}_E = -\lambda\mathbf{M}_A$. M_A and M_B then find equilibrium positions at angles ϕ to z, $-z$, and assuming, as is normally the case in practice, that $H_a \ll H_E$, and that $H \ll H_E$ so that ϕ is small, the equilibrium condition is

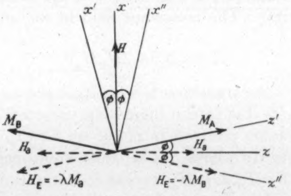


Figure 9.1. Diagram of molecular field model of two-sub-lattice antiferromagnetic, with applied field H perpendicular to easy axis z. M_A, M_B = magnetic moments of sub-lattices ; H_E = antiferromagnetic exchange field ; H_a = anisotropy field ; z', z'' are the equilibrium axes for M_A, $-M_B$ respectively.

$$\phi = H/2H_E. \qquad \ldots\ldots(9.3)$$

The Kittel equations of motion for the system neglecting demagnetizing factors and damping terms are (cf. eqn (6.4))

$$\left.\begin{aligned}\dot{\mathbf{M}}_A &= \gamma\mathbf{M}_A \times (\mathbf{H} - \mathbf{M}_B + \mathbf{H}_a),\\ \dot{\mathbf{M}}_B &= \gamma\mathbf{M}_B \times (\mathbf{H} - \mathbf{M}_A + \mathbf{H}_a),\end{aligned}\right. \qquad \ldots\ldots(9.4)$$

where $\gamma = g\beta/\hbar$ is the gyromagnetic ratio. M_A and M_B are precessing with angular frequency ω about their equilibrium axes z', z'', and we take the small oscillating components of magnetization along the respective perpendicular axes x', y, and x'', y, to be of the form $M_A^{x'} = A\cos\omega t$, $M_A^{y} = A\sin\omega t$, and $M_B^{x''} = B\cos\omega t$, $M_B^{y} = B\sin\omega t$. Then $\dot{M}_A^{y} = \omega M_A^{x'}$, $\dot{M}_B^{y} = \omega M_B^{x''}$, etc. Of the six component equations of (9.4), the pair containing the small quantities $\dot{M}_A^{z'}$, $\dot{M}_B^{z''}$ can be neglected, and the pair containing $\dot{M}_A^{y}$, $\dot{M}_B^{y}$ can be solved separately from those containing $\dot{M}_A^{x'}$, $\dot{M}_B^{x''}$. Those with $\dot{M}_A^{y}$, $\dot{M}_B^{y}$ can be written

$$\left.\begin{aligned}M_A^{x'}\left(\frac{\omega}{\gamma} + H\sin\phi + H_a\cos\phi + H_E\cos 2\phi\right) + M_B^{x''}(H_E\cos 2\phi) = 0\\ M_B^{x''}\left(\frac{\omega}{\gamma} - H\sin\phi - H_a\cos\phi - H_E\cos 2\phi\right) - M_A^{x'}(H_E\cos 2\phi) = 0\end{aligned}\right\}. \quad \ldots\ldots(9.5)$$

Assuming $H_a \ll H_E$ and $H \ll H_E$, equations (9.3) and (9.5) give

$$\left(\frac{\omega}{\gamma}\right)^2 = H^2 + 2H_E H_a. \qquad \ldots\ldots(9.6)$$

There can then be resonant absorption if a small radio-frequency field with angular frequency ω is applied in the usual y direction. Since the usual experimental arrangement is to use a fixed microwave frequency and a variable magnetic field, we put $\omega/\gamma = \hbar\omega/g\beta = H_0$, so that H_0 is the resonance field in the paramagnetic state at temperatures $T > T_N$. If $g = 2$, $H_0 = 10{\cdot}7/\lambda$ kOe where λ is the wavelength in centimetres. The resonance field in the antiferromagnetic state is then, from (9.6),

$$H = (H_0{}^2 - 2H_E H_a)^{\frac{1}{2}}. \qquad \ldots\ldots(9.7)$$

As T is lowered through the transition temperature the resonance line thus shifts from H_0 to lower fields H due to the increasing value of $2H_E H_a$. This effect is observed in the experiments referred to above on $CuCl_2 2H_2O$, Cr_2O_3 (see figure 9.3) and CuMn and AgMn alloys. The shift continues to increase until, at very low temperatures $T \ll T_N$, $2H_E H_a$ reaches its saturation value. If the microwave frequency is not high enough there will be some temperature at which $H_0 = (2H_E H_a)^{1/2}$ in (9.7); the resonance line will then occur in zero field and will not be seen again at lower temperatures.

If we had solved the two component equations of (9.4) which contain $\dot{M}_A{}^{x'}, \dot{M}_B{}^{x''}$, we would have found the condition

$$\left(\frac{\omega}{\gamma}\right)^2 = 2H_E H_a. \qquad \ldots\ldots(9.8)$$

This would only give rise to resonance absorption if the radio-frequency field were directed along x, that is, parallel to the static field H.

9.2.2. *H parallel to easy axis z.*

Provided that the applied field H is not too large, M_A, M_B, are directed along z, $-z$, respectively, and the resonance conditions are found in the same way as before from equation (9.4). However, unlike the perpendicular case, M_A is now slightly larger than M_B except at $T = 0$ when the magnetic moments are saturated. This difference can be represented in terms of measurable quantities by introducing the factor†

$$\alpha = 1 - (\chi_\parallel / \chi_\perp). \qquad \ldots\ldots(9.9)$$

Here, $\chi_\parallel$ is the parallel susceptibility $(M_A - M_B)/H$ which tends to 0 as $T \to 0$; $\chi_\perp$ is the perpendicular susceptibility $(M_A + M_B)\sin\phi/H \simeq M_A/H_E \simeq 1/\lambda \simeq$ constant. It will be noted that $\alpha = 0$ at $T = T_N$, and increases to $\alpha = 1$ at $T = 0$. Also, it can be seen from the expressions for the magnetic energies of the system with H in the parallel and perpendicular positions ($E_\parallel = -\frac{1}{2}H^2\chi_\parallel - \frac{1}{2}(M_A H_a + M_B H_a)$ and $E_\perp = -\frac{1}{2}H^2\chi_\perp$) that since $\chi_\perp > \chi_\parallel$, $E_\perp$ will become less than $E_\parallel$ if $H > H_c$ where

$$H_c{}^2 = \frac{2H_E H_a}{\alpha}. \qquad \ldots\ldots(9.10)$$

---

† In some publications the symbol α denotes $\chi_\parallel/\chi_\perp$.

This means that when H is directed along the easy axis z, and increased through the critical value H_c, M_A and M_B will flop over to be perpendicular to H, and the susceptibility will suddenly increase from $\chi_\parallel$ to $\chi_\perp$, in order to lower the energy. This effect has been observed by the Leiden group in $CuCl_2 2H_2O$; some values of H_c for this salt are shown by curve g–f in figure 9.2.

The resonance conditions found from equation (9.4) will now depend on whether H is greater or less than H_c. We will quote the various solutions below, where it is assumed in all cases that $H_a \ll H_E$.

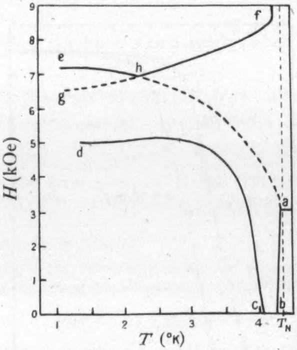


Figure 9.2. Resonance diagram for $CuCl_2 2H_2O$ at 3 cm wavelength, $H_0 \simeq 3000$ Oe, with applied field H along crystal a-axis, which is the easy or z-axis of magnetization (after Ubbink 1953, and Nagamiya, Yosida and Kubo 1955). The experimental points lie approximately on the full curves. Though the symmetry of this salt is actually orthorhombic, the resonance diagram can be used to illustrate the resonance conditions for axial symmetry, with H parallel to z, given in the text. For $T > T_N = 4{\cdot}3$°K, the resonance field $H = H_0$ (paramagnetic region) ; for $T < T_N$, curve abcd $\simeq$ equation (9.11) with the high-temperature end of ba $\simeq$ equation (9.12), and the low-temperature end of cd $\simeq$ equation (9.13) ; ahe $\simeq$ equation (9.16) ; fhg $\simeq$ equation (9.10) $\simeq H_c$ the critical field. Resonance diagrams comparing both 3 cm and 8 to 9 mm measurements with the theory for orthorhombic symmetry are given by Gerritsen (1955).

(*a*) $H \parallel z,\ H < H_c$

This is the normal case, since for the majority of better-known antiferromagnetics where $T_N \sim 100$°K or more, H_c is expected to be much greater than normal laboratory fields of say 10 kOe. The resonance field H is given by

$$H_0^2 = \{\tfrac{1}{2}H(1+\alpha) \pm [2H_E H_a + (1-\alpha)^2(\tfrac{1}{2}H)^2]^{\frac{1}{2}}\}^2. \qquad \ldots\ldots(9.11)$$

We simplify this expression by considering two extreme cases, namely $T \simeq T_N$ and $T \ll T_N$. At $T \simeq T_N$, $\alpha \simeq 0$, and we assume a small temperature region in which $2H_E H_a \ll H^2$. Then (9.11) becomes (using the positive sign)

$$H \simeq H_0 - (2H_E H_a/H), \quad (T \simeq T_N). \qquad \ldots\ldots(9.12)$$

Though (9.12) is not likely to be valid over more than an extremely small temperature range, it indicates that as T is lowered through T_N the resonance line is expected to shift from H_0 to lower field H due to the increasing value of $2H_E H_a$.

This is found experimentally in $CuCl_2 2H_2O$ (curve a–b in figure 9.2) and in Cr_2O_3 (figure 9.3). The line shifts rapidly to zero field and disappears, and at lower temperatures reappears when (9.11) is satisfied with the negative sign (curve c–d in figure 9.2). When $T \ll T_N$, $(1-\alpha) \ll 1$, $(2H_E H_a)^{\frac{1}{2}}$ is nearly constant, and (9.11) becomes

$$H(1+\alpha)/2 \simeq |(2H_E H_a)^{\frac{1}{2}} - H_0| \quad (T \ll T_N). \qquad \ldots\ldots(9.13)$$

Normally, $(2H_E H_a)^{\frac{1}{2}}$ is expected to be greater than H_0, and the modulus signs can then be omitted. (In any case solutions with $H > (2H_E H_a)^{\frac{1}{2}}$, $\simeq H_c$ since $\alpha \simeq 1$, are not allowed, since (9.11) is derived on the assumption that $H < H_c$.) Equation (9.13) gives a resonance field H which varies only slowly with temperature, and for $CuCl_2 2H_2O$ at 3 cm wavelength it corresponds to the low-temperature end of curve c–d in figure 9.2. This resonance condition (9.13) for H parallel to z, at $T=0$ where $(1+\alpha)/2 = 1$, and with $H_a \ll H_E$, is the one originally derived by Kittel (1951).

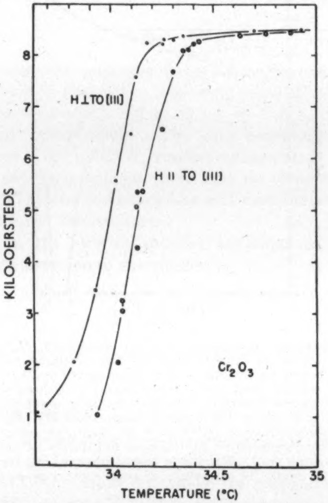


Figure 9.3. Resonance diagram for Cr_2O_3 at 1·2 cm wavelength with applied field H parallel and perpendicular to (1, 1, 1) which is the easy or z-axis of magnetization (*after* McGuire, Scott and Grannis 1956). For $T > T_N = 34{\cdot}5$°c, $H = H_0$ (paramagnetic region); for $T < T_N$, the $H_\perp$ resonance field follows equation (9.7); the $H_\parallel$ field follows equation (9.11), and, for a small temperature region close to T_N, equation (9.12). The relation between $H_\parallel$ and $H_\perp$ follows equation (9.15).

Finally, we write equation (9.11) in the form which is convenient if the microwave frequency ($\nu = \gamma H_0/2\pi$) is being varied, rather than the applied field H.

If $H \ll (2H_E H_a)^{\frac{1}{2}}$, this is

$$\nu = \frac{\gamma}{2\pi}[(2H_E H_a)^{\frac{1}{2}} \pm \tfrac{1}{2}H(1+\alpha)].$$

Resonance absorption occurs in zero field ($H = 0$) if an appropriate value of $(2H_E H_a)^{\frac{1}{2}}$ can be found by varying the temperature. Johnson and Nethercot (1956) have made experiments of this sort on MnF_2 with $T = 42$ to $64°\text{K}$, where $T_N = 67{\cdot}7°\text{K}$. They find $\nu_{H=0} = 264 B_{5/2}(T/T_N) \times 10^9$ c/s, where $B_{5/2}$ is a modified Brillouin function for $S = 5/2$.

(*b*) $H < H_c$, $H\hat{O}z = \theta$

The general resonance condition which applies when H is directed at angle θ to z is

$$H_0{}^2 = 2H_E H_a + \tfrac{1}{2}H^2[1+\alpha^2\cos^2\theta] \pm \tfrac{1}{2}H\{8H_E H_a(1+\alpha)^2\cos^2\theta + H^2[\sin^4\theta + \cos^2\theta(1+\alpha)^2(2\sin^2\theta + (1-\alpha)^2\cos^2\theta)]\}^{\frac{1}{2}}. \qquad \ldots\ldots(9.14)$$

This is expressed in the form given by Keffer (1952), except for our different definition of α, equation (9.9). It reduces to (9.11) if $\theta = 0$, and to (9.7) if $\theta = \pi/2$.

(*c*) $H < H_c$, *relation between* $H_\parallel$ *and* $H_\perp$

The relation between the resonance fields for the parallel and perpendicular positions of H, which is found by eliminating $2H_E H_a$ from (9.11) and (9.7), is

$$[\alpha H_\parallel{}^2 + H_\perp{}^2]/H_\parallel H_0 = 1 + \alpha. \qquad \ldots\ldots(9.15)$$

This may be useful for comparing experiment with theory when the value of $2H_E H_a$ is unknown. It has been used by McGuire, Scott and Grannis (1956) (see also Wangsness 1956) on their Cr_2O_3 results (figure 9.3) in the region close to T_N where $2H_E H_a$ is changing very rapidly with temperature. They find reasonable agreement with (9.15), and in particular, very close to T_N where $\alpha \simeq 0$ the relation $H_\perp{}^2 \simeq H_\parallel H_0$ is satisfied. It will be noted that this relation suggests, in agreement with observation, that $H_\parallel$ is always less than or equal to $H_\perp$ in this region.

(*d*) $H > H_c$, $H \parallel z$

This case only arises in practice if the critical field is small, which usually implies a low transition temperature, T_N. The resonance field is

$$H = (H_0{}^2 + 2H_E H_a)^{\frac{1}{2}}. \qquad \ldots\ldots(9.16)$$

The field corresponding to (9.16) in $CuCl_2 2H_2O$ at 3 cm wavelength is a–e in figure 9.2. This lies above the critical field g–f only at the low-temperature end h–e, so that resonance corresponding to (9.16) is only seen along h–e.

(*e*) $H = H_c$

Under certain conditions (Nagamiya 1954) there can also be resonance absorption if the applied field is exactly equal to the critical field (eqn (9.10)) ; this has been called orientation resonance. It has been found by Ubbink (1953) along the region f–h of the critical field curve f–g in figure 9.2. For further details reference should be made to the various publications quoted above.

9.2.3. *Intensity of absorption.*

The first simple estimates of intensity were made by Keffer and Kittel (1952), who showed that with $H \parallel z$, the ratio of the radio-frequency susceptibilities at 0°K to just above the transition point T_N is of the order of $(H_E H_a)^{\frac{1}{2}}/H$. Thus, even though the static susceptibility $\chi_{\parallel}$ is zero at 0°K, the absorption intensity may be quite high. For $CuCl_2 2H_2O$ at 1°K for instance, the intensity might be expected to be of the same order as it would be if the salt were paramagnetic at 1°K. The only detailed investigation is that by Gerritsen and Garber (1956) in $CuCl_2 2H_2O$. They find the absolute magnitude of the intensity to be of the order mentioned above, and the variation with the direction of the applied static field and radio-frequency field to be in good agreement with their detailed theory.

9.2.4. *Line widths.*

Some of the possible contributions to line width include (i) small unresolved splittings of the resonance line arising from demagnetizing effects (Keffer and Kittel 1952); (ii) broadening due to crystal imperfections causing local anisotropy fields; (iii) exchange narrowing. The most detailed experimental investigation has been by Gerritsen, Garber and Drewes (1956) in $CuCl_2 2H_2O$ at 3 cm wavelength. They find a complicated dependence of line width on temperature which cannot be accounted for by any simple explanation. For the resonance at fields c–d in figure 9.2 the width (i.e. distance between the half-power points) increases rapidly with temperature, being 20 Oe at 1·3°K, 140 Oe at 3°K, and about 3000 gauss at 4·1°K when the line occurs in zero field. The other absorption seen along e–h–f has a minimum width of 10 to 30 Oe from 2·5 to 3·4°K. These values compare with a width of about 80 Oe in the paramagnetic region at 20°K. Gerritsen *et al.* suggest that the large width found for lines occurring near zero field may be partly because in this region a small frequency width is equivalent to a large magnetic field width. The antiferromagnetic resonance lines in Cr_2O_3 referred to above are also characterized by a width which increases to several thousand Oe as the lines move towards very low fields. In MnF_2, where the experiments were designed to observe all the resonance lines in or close to zero field, Johnson and Nethercot (1956) find that the width is 8000 Oe at $T = 63{\cdot}6$°K, decreasing steadily to 1600 Oe at 42·0°K.

ACKNOWLEDGMENTS

We would like to thank Drs. R. J. Elliott, J. H. E. Griffiths and B. R. Judd for helpful discussions and suggestions. We are also grateful to the authors and publishers of papers from which diagrams have been reproduced for permission to reproduce them.

REFERENCES

§ 1. INTRODUCTION

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London : Physical Society).

Bowers, K. D., and Owen, J., 1955, *Rep. Progr. Phys.*, **18**, 305 (London : Physical Society).

Hutchison, C. A., 1956, *Annu. Rev. Phys. Chem.*, **7**, 359.

Ingram, D. J. E., 1955, *Spectroscopy at Radio and Microwave Frequencies* (London : Butterworths Scientific Publications).

Overhauser, A. W., 1953, *Phys. Rev.*, **91**, 476 ; **92**, 411.

Wertz, J. E., 1955, *Chem. Rev.*, **55**, 829.

§ 2. DEFECTS IN CRYSTALS

Ard, W. B., 1955, *J. Chem. Phys.*, **23**, 1967.

Ard, W. B., and Gordy, W., 1955, *Phys. Rev.*, **99**, 612.

Ard, W. B., Shields, H., and Gordy, W., 1955, *J. Chem. Phys.*, **23**, 1727.

Beringer, R., and Heald, M. A., 1954, *Phys. Rev.*, **95**, 1474.

Bleaney, B., 1956, *Phys. Rev.*, **104**, 1190.

Bleaney, B., Hayes, W., and Llewellyn, P. M., 1957, *Nature, Lond.*, **179**, 140.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London : Physical Society).

Bowers, K. D., and Owen, J., 1955, *Rep. Progr. Phys.*, **18**, 304 (London : Physical Society).

Burgess, J. H., and Norberg, R. E., 1955, *Phys. Rev.*, **100**, 752.

Castle, J. G., and Beringer, R., 1950, *Phys. Rev.*, **80**, 114.

Cohen, M. H., 1956, *Phys. Rev.*, **101**, 1432.

Combrisson, J., and Uebersfeld, J., 1954, *C.R. Acad. Sci., Paris*, **238**, 1397.

Delbeçq, C. J., Smaller, B., and Yuster, P. H., 1956, *Phys. Rev.*, **104**, 599.

Dexter, D. L., 1954, *Phys. Rev.*, **93**, 244.

Feher, G., 1956, *Phys. Rev.*, **103**, 834.

Feher, G., and Fletcher, R. C., 1956, *Bull. Amer. Phys. Soc.*, **1**, Ser. 2, 125.

Feher, G., Fletcher, R. C., and Gere, E. A., 1955, *Phys. Rev.*, **100**, 1784.

Fletcher, R. C., Yager, W. A., Pearson, G. L., Holden, A. N., Read, W. T., and Merritt, F. R., 1954, *Phys. Rev.*, **94**, 1392.

Fletcher, R. C., Yager, W. A., Pearson, G. L., and Merritt, F. R., 1954, *Phys. Rev.*, **95**, 844.

Gordy, W., Ard, W. B., and Shields, H., 1956, *Proc. Nat. Acad. Sci., Wash.*, **41**, 983, 996.

Griffiths, J. H. E., Owen, J., and Ward, I. M., 1955, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society), p. 81.

Honig, A., and Kip, A. F., 1954, *Phys. Rev.*, **95**, 1686.

Hutchison, C. A., 1949, *Phys. Rev.*, **75**, 1769.

Hutchison, C. A., and Noble, G. A., 1952, *Phys. Rev.*, **87**, 1125.

Inui, T., Harasawa, S., and Obata, Y., 1956, *J. Phys. Soc., Japan*, **11**, 612.

Inui, T., and Uemura, Y., 1950, *Progr. Theor. Phys., Japan*, **5**, 252, 395.

Jen, C. K., Foner, S. N., Cochran, E. L., and Bowers, V. A., 1956, *Phys. Rev.*, **104**, 846.

Kahn, A. H., and Kittel, C., 1953, *Phys. Rev.*, **89**, 315.

Kanzig, W., 1955, *Phys. Rev.*, **99**, 1890.

Kip, A. F., Kittel, C., Levy, R. A., and Portis, A. M., 1953, *Phys. Rev.*, **91**, 1066.

Kittel, C., 1955, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society), p. 33.

Kittel, C., and Mitchell, A., 1954, *Phys. Rev.*, **96**, 1488.

Kohn, W., and Luttinger, J. M., 1955, *Phys. Rev.*, **97**, 883.

Krumhansl, J. A., 1954, *Phys. Rev.*, **93**, 245.

Kusch, P., 1955, *Phys. Rev.*, **100**, 1188.

Lark-Horowitz, K., Schulz-DuBois, E., Nisenoff, M., and Fan, H. Y., 1955, *Phys. Rev.*, **98**, 1561.

Livingston, R., Zeldes, H., and Taylor, E. H., 1955, *Disc. Faraday Soc.*, No. 19, p. 166.

LOW, W., 1956, *Phys. Rev.*, **101**, 1827.

LUTTINGER, J. M., and KOHN, W., 1955, *Phys. Rev.*, **97**, 869.

MATHESON, M. S., and SMALLER, B., 1955, *J. Chem. Phys.*, **23**, 521.

MUTO, T., 1949, *Progr. Theor. Phys., Japan*, **4**, 243.

O'BRIEN, M. C. M., 1955, *Proc. Roy. Soc.* A, **231**, 404.

O'BRIEN, M. C. M., and PRYCE, M. H. L., 1955, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society), p. 88.

PORTIS, A. M., 1953, *Phys. Rev.*, **91**, 1071.

PORTIS, A. M., and SHALTIEL, D., 1955, *Phys. Rev.*, **98**, 264.

SANDS, R. H., TOWNSEND, J., and HOOD, H. P., 1954, *Phys. Rev.*, **95**, 607.

SCHNEIDER, E. E., 1954, *Phys. Rev.*, **93**, 919 ; 1955, *Disc. Faraday Soc.*, No. 19, p. 158.

SCHNEIDER, E. E., DAY, M. J., and STEIN, G., 1951, *Nature, Lond.*, **168**, 645.

SCHNEIDER, E. E., and ENGLAND, T. S., 1951, *Physica, 's Grav.*, **17**, 221.

SEITZ, F., 1954, *Rev. Mod. Phys.*, **26**, 7.

SHULMAN, R. G., and JACCARINO, V., 1956, *Phys. Rev.*, **103**, 1126.

SIMPSON, J. H., 1949, *Proc. Roy. Soc.* A, **197**, 269.

SLICHTER, C. P., 1955, *Phys. Rev.*, **99**, 479.

SMALLER, B., DELBECQ, C. J., and YUSTER, P. H., 1956, *Bull. Amer. Phys. Soc.*, **1**, 324.

SMALLER, B., MATHESON, M. S., and YASAITIS, E. L., 1954, *Phys. Rev.*, **94**, 202.

TINKHAM, M., 1956, *Proc. Roy. Soc.* A, **236**, 535, 549.

TINKHAM, M., and KIP, A. F., 1951, *Phys. Rev.*, **83**, 657.

WERTZ, J. E., and VIVO, J. L., 1955, *Phys. Rev.*, **100**, 1792.

YASAITIS, E. L., and SMALLER, B., 1953, *Phys. Rev.*, **92**, 1068.

§ 3. CONDUCTORS

AL'TSHULER, S. A., KURENEV, V. YA., and SALIKHOV, S. G., 1952, *Doklady Akad. Nauk. S.S.S.R.*, **84**, 677.

ARGYRES, P. N., and KAHN, A. H., *Abstract for 1954 Chicago meeting of the American Physical Society*.

BIONDI, M. A., GARFUNKEL, M. P., and MCCOUBREY, A. O., 1956, *Phys. Rev.*, **101**, 1427.

BLEVINS, G. S., GORDY, W., and FAIRBANK, W. M., 1955, *Phys. Rev.*, **100**, 1215.

BLOEMBERGEN, N., 1952, *J. Appl. Phys.*, **23**, 1379.

BROOKS, H., 1954, *Phys. Rev.*, **94**, 1411.

BUCKINGHAM, M. J., 1956, *Phys. Rev.*, **101**, 1431.

CHAMBERS, R. G., 1952, *Proc. Roy. Soc.* A, **215**, 481.

DYSON, F. J., 1955, *Phys. Rev.*, **98**, 349.

ELLIOTT, R. J., 1954, *Phys. Rev.*, **96**, 266, 280.

FEHER, G., and KIP, A. F., 1955, *Phys. Rev.*, **98**, 337.

FRÖHLICH, F., and NABARRO, F. R. N., 1940, *Proc. Roy. Soc.* A, **175**, 382.

GLOVER, R. E., and TINKHAM, M., 1956, *Phys. Rev.*, **104**, 844.

GRISWOLD, T. W., KIP, A. F., and KITTEL, C., 1952, *Phys. Rev.*, **88**, 951.

GUISBURG, V. L., and LANDAU, L. D., 1950, *J. Exp. Theor. Phys.*, **20**, 1064.

GUTOWSKY, H. S., and FRANK, P. J., 1954, *Phys. Rev.*, **95**, 1067.

KIP, A. F., 1955, *Report of the Conference on Defects in Crystalline Solids* (London : Physical Society).

KITTEL, C., 1954, *Proceedings of the 10th Solvay Conference, Brussels* (Brussels : Institut de Sociologie Solvay).

KITTEL, C., and MITCHELL, A. H., 1956, *Phys. Rev.*, **101**, 1611.

LEVY, R. A., 1956, *Phys. Rev.*, **102**, 31.

LONDON, H., 1940, *Proc. Roy. Soc.* A, **176**, 522.

OVERHAUSER, A. W., 1953, *Phys. Rev.*, **89**, 689.

OWEN J., BROWNE, M. E., ARP, V., and KIP, A. F., 1957, *J. Phys. Chem. Solids*, **2**, 85.

OWEN, J., BROWNE, M. E., KNIGHT, W. D., and KITTEL, C., 1956, *Phys. Rev.*, **102**, 1501.

PINES, D., 1954, *Phys. Rev.*, **95**, 1090.

PINES, D., and BOHM, D., 1952, *Phys. Rev.*, **85**, 338.

PIPPARD, A. B., 1947, *Proc. Roy. Soc.* A, **191**, 385 ; 1950, *Ibid.*, **203**, 98 ; 1954, *Electronics and Electron Physics* (New York : Academic Press).

REUTER, G., and SONDHEIMER, E., 1948, *Proc. Roy. Soc.* A, **195**, 336.

SALIKHOV, S. G., 1952, *Akad. Nauk. S.S.S.R.*, No. 6, 748.

SCHUMACHER, R. T., CARVER, T. R., and SLICHTER, C. P., 1954, *Phys. Rev.*, **95**, 1089.

SCHUMACHER, R. T., and SLICHTER, C. P., 1956, *Phys. Rev.*, **101**, 58.

SOLT, I. H., and STRANDBERG, M. W. P., 1954, *Phys. Rev.*, **95**, 607.

SONDHEIMER, E. H., 1952, *Advanc. Phys.* (*Phil. Mag. Suppl.*), **1**, 1.

TINKHAM, M., 1956, *Phys. Rev.*, **104**, 844.

YAFET, Y., 1952, *Phys. Rev.*, **85**, 478.

ZENER, C., 1951, *Phys. Rev.*, **81**, 440.

§ 4. ELECTRICAL ABSORPTION IN SEMICONDUCTORS

ARTHUR, J. B., GIBSON, A. F., and GRANVILLE, J. W., 1956, *J. Electron.*, **2**, 145.

AZBEL, M. YA., and KANER, E. A., 1956, *J. Exp. Theor. Phys.*, **30**, 811.

BAGGULEY, D. M. S., POWELL, J. A., and TAYLOR, D. J., 1957 (to be published).

BENEDICT, T. S., 1953, *Phys. Rev.*, **91**, 1565.

BENEDICT, T. S., and SHOCKLEY, W., 1953, *Phys. Rev.*, **89**, 1152.

BURSTEIN, E., PICUS, G. S., and GEBBIE, H. A., 1956, *Phys. Rev.*, **103**, 825.

D'ALTROY, F. A., and FAN, H. Y., 1954, *Phys. Rev.*, **94**, 1415.

DEXTER, R. N., and LAX, B., 1955 a, *Phys. Rev.*, **100**, 1216 ; 1955 b, *Ibid.*, **99**, 635.

DEXTER, R. N., ZEIGER, H. J., and LAX, B., 1956, *Phys. Rev.*, **104**, 637.

DINGLE, R. B., 1952, *Proc. Roy. Soc.* A, **212**, 38.

DORFMAN, J. G., 1951, *Doklady Akad. Nauk. S.S.S.R.*, **81**, 765.

DRESSELHAUS, G., KIP, A. F., HAN-YING KU, WAGONER, G., and CHRISTIAN, S. M., 1955, *Phys. Rev.*, **100**, 1218.

DRESSELHAUS, G., KIP, A. F., and KITTEL, C., 1955 a, *Phys. Rev.*, **98**, 368 ; 1955 b, *Ibid.*, **98**, 556 ; 1955 c, *Ibid.*, **100**, 618.

FAN, H. Y., 1955, *Solid State Physics*, Vol. I (Academic Press).

FAN, H. Y., and BECKER, N., 1951, *Semi-conducting Materials*, Ed. H. K. Henisch (Butterworths Scientific Publications).

FAWCETT, E., 1956, *Phys. Rev.*, **103**, 1582.

FLETCHER, R. C., YAGER, W. A., and MERRITT, F. R., 1955, *Phys. Rev.*, **100**, 747.

GALT, J. K., YAGER, W. A., MERRITT, F. R, CETLIN, B. B., and DAIL, H. W., 1955, *Phys. Rev.*, **100**, 748.

GIBSON, A. F., 1956, *Proc. Phys. Soc.* B, **69**, 488.

GIBSON, A. F., and GRANVILLE, J. W., 1956, *J. Electron.*, **2**, 259.

GUNN, J. B., and HOGARTH, C. A., 1955, *J. Appl. Phys.*, **26**, 353.

LAX, B., BUTTON, K. J., ZEIGER, H. J., and ROTH, L. M., 1956, *Phys. Rev.*, **102**, 715.

LAX, B., ZEIGER, H. J., and DEXTER, R. N., 1954, *Physica, 's Grav.*, **20**, 818.

LUTTINGER, J. M., and KOHN, W., 1954, *Phys. Rev.*, **96**, 529 ; 1955, *Ibid.*, **97**, 869.

SEITZ, F., 1940, *Modern Theory of Solids* (New York : McGraw-Hill).

SHOCKLEY, W., 1953, *Phys. Rev.*, **90**, 491.

TINKHAM, M., 1956, *Phys. Rev.*, **101**, 902.

ZWERDLING, S., KEYES, R. J., FONER, S., KOLM, H. H., LIPSOM, H., WARSCHAUER, D., and LAX, B., 1956, *Bull. Amer. Phys. Soc.*, Ser. 2, **1**, 299.

§ 5. SPIN RESONANCE FROM PAIRS OF COUPLED ATOMS

ABE, H., and SHIMADA, J., 1953, *Phys. Rev.*, **90**, 316.

ABE, H., ONO, K., HAYASHI, I., SHIMADA, J., and IWANAGA, K., 1954, *J. Phys. Soc. Japan*, **9**, 814.

ANDERSON, P. W., 1950, *Phys. Rev.*, **79**, 705 ; 1954, *J. Phys. Soc. Japan*, **9**, 216.

ANDERSON, P. W., and WEISS, M. T., 1953, *Rev. Mod. Phys.*, **25**, 269.

BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1948, *Nature, Lond.*, **162**, 538.

Baker, J. M., and Bleaney, B., 1955, *Conference on the Physics of Low Temperatures*, 1955 (Paris : Institut International du Froid).

Bleaney, B., and Bowers, K. D., 1952, *Proc. Roy. Soc.* A, **214**, 451.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London : Physical Society).

Cooke, A. H., McKim, F. R., Smith, P. L., and Wolf, W. P., 1957 (to be published).

Feher, G., Fletcher, R. C., and Gere, E. A., 1955, *Phys. Rev.*, **100**, 1784.

Fletcher, R. C., Yager, W. A., Pearson, G. L., and Merritt, F. R., 1954, *Phys. Rev.*, **95**, 844.

Griffiths, J. H. E., Owen, J., Park, G., and Partridge, M. F., 1957 (to be published).

Guha, B. C., 1951, *Proc. Roy. Soc.* A, **206**, 353.

Kramers, H. A., 1930, *Proc. Acad. Sci. Amst.*, **33**, 959 ; 1934, *Physica*, **1**, 182.

Ollom, J. F., and Van Vleck, J. H., 1951, *Physica, 's Grav.*, **17**, 205.

Opechowski, W., 1948, *Physica, 's Grav.*, **14**, 237.

Pryce, M. H. L., 1948, *Nature, Lond.*, **162**, 539.

Schneider, E. E., 1955, *Disc. Faraday Soc.*, No. 19, 158.

Slichter, C. P., 1955, *Phys. Rev.*, **99**, 479.

Stevens, K. W. H., 1953, *Rev. Mod. Phys.*, **25**, 166.

van Niekerk, J. N., and Schoening, F. R. L., 1953, *Acta Cryst., Camb.*, **6**, 227.

Van Vleck, J. H., 1937, *Phys. Rev.*, **52**, 1178 ; 1941, *J. Chem. Phys.*, **9**, 85 ; 1948, *Phys. Rev.*, **74**, 1168.

van Wieringen, J. S., 1955, *Disc. Faraday Soc.*, No. 19, 118.

Yosida, K., 1950, *Progr. Theor. Phys., Japan*, **5**, 1047.

§ 6. General Discussion of Extensive Coupled Systems

Bickford, L. R., 1950, *Phys. Rev.*, **78**, 449.

Bloch, F., 1946, *Phys. Rev.*, **70**, 460.

Bloembergen, N., 1950, *Phys. Rev.*, **78**, 572.

Dillon, J. F., Geschwind, S., and Jaccarino, V., 1955, *Phys. Rev.*, **100**, 750.

Kittel, C., 1948, *Phys. Rev.*, **73**, 155 ; 1949, *Ibid.*, **76**, 743.

Landau, L., and Lifshitz, E., 1935, *Physik. Z. Sowjet.*, **8**, 153.

Néel, L., 1948, *Ann. Phys., Paris*, **3**, 137.

Polder, D., 1949, *Phil. Mag.*, **40**, 99.

Smit, J., and Wijn, H. P. J., 1954, *Electronics and Electron Physics, VI* (New York : Academic Press).

Van Vleck, J. H., 1950, *Phys. Rev.*. **78**, 266.

Wangsness, R. K., 1953, *Phys. Rev.*, **91**, 1085.

Yager, W. A., Galt, J. K., Merritt, F. R., and Wood, E. A., 1950, *Phys. Rev.*, **80**, 744.

§ 7. Ferromagnetic Metals

Abrahams, E., 1954, *Electronics and Electron Physics, VI* (New York : Academic Press).

Bagguley, D. M. S., 1953, *Proc. Phys. Soc.* A, **66**, 765 ; 1955, *Ibid.*, **228**, 549.

Bagguley, D. M. S., and Harrick, N. J., 1954, *Proc. Phys. Soc.* A, **67**, 648.

Barlow, G. S., and Standley, K. J., 1956, *Proc. Phys. Soc.* B, **69**, 1052.

Barnett, S. J., 1944, *Proc. Amer. Acad. Arts Sci.*, **75**, 109.

Bloembergen, N., 1950, *Phys. Rev.*, **78**, 572.

Bozorth, R. M., 1951, *Ferromagnetism* (New York : van Nostrand).

Brown, S., Meyer, A. J. P., and Scott, G. G., 1954, *C.R. Acad. Sci., Paris*, **238**, 2504.

Clogston, A. M., Suhl, H., Walker, L. R., and Anderson, P. W., 1956, *Phys. Rev.*, **101**, 903.

Fletcher, G. C., 1954, *Proc. Phys. Soc.* A, **67**, 505.

Griffiths, J. H. E., 1946, *Nature, Lond.*, **158**, 670 ; 1951, *Physica, 's Grav.*, **17**, 253.

Griffiths, J. H. E., and MacDonald, J. R., 1951, *J. Sci. Instrum.*, **28**, 56.

Hoskins, R., and Weiner, G., 1954, *Phys. Rev.*, **96**, 1153.

Kip, A. F., 1953, *Rev. Mod. Phys.*, **25**, 229.

Kip, A. F., and Arnold, R. D., 1948, *R.L.E. Prog. Rep. M.I.T.*, Oct. 15, p. 31 ; 1949, *Phys. Rev.*, **75**, 1556.
Kittel, C., 1948, *Phys. Rev.*, **73**, 155.
Kittel, C., and Abrahams, E., 1953, *Rev. Mod. Phys.*, **25**, 233.
Kittel, C., and Mitchell, A. H., 1956, *Phys. Rev.*, **101**, 1611.
Meyer, A. J. P., 1952, *C.R. Acad. Sci., Paris*, **235**, 1382.
Reich, K. H., 1956, *Phys. Rev.*, **101**, 1647.
Scott, G. G., 1955 a, *Phys. Rev.*, **99**, 1241 ; 1955 b, *Ibid.*, **99**, 1824.
Standley, K. J., 1949, *Thesis*, University of Oxford.
Standley, K. J., and Reich, K. H., 1955, *Proc. Phys. Soc.* B, **68**, 713.
Standley, K. J., and Stevens, K. W. H., 1956, *Proc. Phys. Soc.* B, **69**, 993.
Van Vleck, J. H., 1950, *Phys. Rev.*, **78**, 266.
Yager, W. A., 1949, *Phys. Rev.*, **75**, 316.
Yager, W. A., and Merritt, F. R., 1949, *Phys. Rev.*, **75**, 318.

§ 8. Ferrite Materials

Abrahams, E., 1954, *Electronics and Electron Physics, VI* (New York : Academic Press).
Anderson, P. W., 1952, *Phys. Rev.*, **88**, 1214.
Becker, R., 1951, *J. Phys. Radium*, **12**, 332.
Beljers, H. G., Van der Lindt, W. J., and Went, J. J., 1951, *J. Appl. Phys.*, **22**, 1506.
Bickford, L. R., 1950, *Phys. Rev.*, **78**, 449.
Bloembergen, N., 1952, *J. Appl. Phys.*, **23**, 1383.
Bloembergen, N., and Wang, S., 1954, *Phys. Rev.*, **93**, 72.
Bowers, K. D., and Owen, J., 1955, *Rep. Progr. Phys.*, **18**, 304.
Damon, R. W., 1953, *Rev. Mod. Phys.*, **25**, 239.
Dillon, J. F., Geschwind, S., and Jaccarino, V., 1955, *Phys. Rev.*, **100**, 750.
Doring, W., 1948, *Z. Naturforsch.*, **3a**, 374.
Fairweather, A., Roberts, F. F., and Welch, A. J. E., 1952, *Rep. Progr. Phys.*, **15**, 142.
Fox, A. G., Miller, S. E., and Weiss, M. T., 1955, *Bell Syst. Tech. J.*, **34**, 5.
Galt, J. K., Andrus, J., and Hopper, H. G., 1953, *Rev. Mod. Phys.*, **25**, 93.
Galt, J. K., Yager, W. A., Remeika, J. P., and Merritt, F. R., 1951, *Phys. Rev.*, **81**, 470.
Hogan, C. L., 1953, *Rev. Mod. Phys.*, **25**, 253.
Johnson, M. H., and Rado, G. T., 1949, *Phys. Rev.*, **75**, 841.
Kittel, C., 1948, *Phys. Rev.*, **73**, 155 ; 1949, *Rev. Mod. Phys.*, **21**, 541 ; 1951, *J. Phys. Radium*, **12**, 291.
Néel, L., 1948, *Ann. Phys., Paris*, **3**, 137.
Okamura, T., and Kojima, Y., 1952, *Phys. Rev.*, **86**, 1040.
Polder, D., 1949, *Phil. Mag.*, **40**, 99.
Polder, D., and Smit, J., 1953, *Rev. Mod. Phys.*, **25**, 89.
Rado, G. T., 1953, *Rev. Mod. Phys.*, **25**, 81.
Rowen, J. H., 1953, *Bell Syst. Tech. J.*, **32**, 1333.
Smit, J., and Wijn, H. P. J., 1954, *Electronics and Electron Physics, VI* (New York : Academic Press).
Snoek, J. L., 1948, *Physica, 's Grav.*, **14**, 207.
Suhl, H., and Walker, L. R., 1954, *Bell Syst. Tech. J.*, **33**, pages (a) 579, (b) 939, (c) 1133.
Tannenwald, P. E., 1955 a, *Phys. Rev.*, **99**, 463 ; 1955 b, *Ibid.*, **100**, 1713.
Weiss, M. T., and Fox, A. G., 1952, *Phys. Rev.*, **88**, 146.
Wijn, H. P. J., 1953, *Physica, 's Grav.*, **19**, 555.
Wijn, H. P. J., Gevas, M., and Van der Burgt, C. M., 1953, *Rev. Mod. Phys.*, **25**, 91.
Yager, W. A., Galt, J. K., and Merritt, F. R., 1955, *Phys. Rev.*, **99**, 1203.
Yager, W. A., Galt, J. K., Merritt, F. R., and Wood, E. A., 1950, *Phys. Rev.*, **80**, 744.

§ 9. Antiferromagnetic Resonance

Date, M., 1956, *Phys. Rev.*, **104**, 623.
Dayhoff, E. S., 1956, *Bull. Amer. Phys. Soc.*, **1**, 283.

Gerritsen, H. J., 1955, *Physica, 's Grav.*, **21**, 639.
Gerritsen, H. J., and Garber, M., 1956, *Physica, 's Grav.*, **22**, 197, 481.
Gerritsen, H. J., Garber, M., and Drewes, G. W. J., 1956, *Physica, 's Grav.*, **22**, 213.
Gerritsen, H. J., Okkes, R. F., Bölger, B., and Gorter, C. J., 1955, *Physica, 's Grav.*, **21**, 629.
Gorter, C. J., and Hantjes, J., 1952, *Physica, 's Grav.*, **18**, 285.
Johnson, F. M., and Nethercott, A. H., 1956, *Phys. Rev.*, **104**, 847.
Keffer, F., 1952, *Phys. Rev.*, **87**, 608.
Keffer, F., and Kittel, C., 1952, *Phys. Rev.*, **85**, 329.
Kittel, C., 1951, *Phys. Rev.*, **82**, 565.
Lidiard, A. B., 1954, *Rep. Progr. Phys.*, **17**, 201 (London : Physical Society).
McGuire, T. R., Scott, E. J., and Grannis, F. H., 1956, *Phys. Rev.*, **102**, 1000.
Maxwell, L. R., and McGuire, T. R., 1953, *Rev. Mod. Phys.*, **25**, 279.
Nagamiya, T., 1951, *Progr. Theor. Phys., Japan*, **6**, 342 ; 1954, *Ibid.*, **11**, 309.
Nagamiya, T., Yosida, K., and Kubo, R., 1955, *Advanc. Phys.*, **4**, 1.
Owen, J., Browne, M. E., Arp, V., and Kip, A. F., 1957, *J. Phys. Chem. Solids*, **2**, 85.
Owen, J., Browne, M. E., Knight, W. D., and Kittel, C., 1956, *Phys. Rev.*, **102**, 1501.
Poulis, N. J., van den Handel, J., Ubbink, J., Poulis, J. A., and Gorter, C. J., 1951, *Phys. Rev.*, **82**, 552.
Trounson, E. P., Bleil, D. F., Wangsness, R. K., and Maxwell, L. R., 1950, *Phys. Rev.*, **79**, 542.
Ubbink, J., 1953, *Thesis*, University of Leiden.
Ubbink, J., Poulis, J. A., Gerritsen, J. H., and Gorter, C. J., 1953, *Physica, 's Grav.*, **19**, 928.
Van Vleck, J. H., 1951, *J. Phys. Radium*, **12**, 262.
Wangsness, R. K., 1956, *Bull. Amer. Phys. Soc.*, **1**, 126.

A REVIEW OF THE METHODS OF HIGH-SPEED PHOTOGRAPHY

By J. S. COURTNEY-PRATT

Research Laboratory for the Physics and Chemistry of Surfaces,
Department of Physical Chemistry, University of Cambridge

CONTENTS

PAGE
§ 1. Introduction 380
§ 2. Streak records 383
§ 3. Single exposures 386
§ 4. Multiple exposure without displacement of the frames 396
§ 5. Series of separate pictures 397
§ 6. Series of pictures by image dissection 411
§ 7. Conclusion 428
References 429

Abstract. The variety, range and precision of methods available for photographic recording of fast phenomena have been increasing rapidly. The capabilities of the newer techniques are considered, classifying them by the kind of record obtained. Streak records with drum cameras can give a time resolution of 10^{-7} sec; rotating mirror cameras at present approach 10^{-8} sec and may eventually achieve $\frac{1}{4} \times 10^{-9}$ sec; deflecting image converters may go much further. Single flashes of light, bright enough for silhouette recording, can be as short as 10^{-8} sec, and for reflected-light recording 10^{-7} sec for near objects, or 10^{-6} sec for a field of view a metre square. Kerr cells can operate with an exposure of 10^{-8} sec, and image converter tubes 10^{-9} sec. Several pictures can easily be taken at short intervals superimposed on the one plate. Frames may be separated by intermittent film movement at rates less than 300 pictures per second (p.p.s.); or for speeds up to 10^4 p.p.s. by using continuously moving film and short exposures, or by using some form of image movement compensation where exposures are a significant proportion of the interframe interval. With smaller pictures of lower resolution, higher speeds are possible. For higher speeds still one uses effectively separate cameras exposed in succession by mechanical means (such as a rotating slotted disc) up to 10^5 p.p.s., by optical means (such as the use of a rotating mirror) up to 10^7 p.p.s., or by electronic means (such as phased shutters or phased spark sources) up to 10^7 p.p.s. Comparably high speeds with less elaborate equipment are made possible by image dissection. Simple dissection plates with clear lines or holes in an opaque ground allow recording rates of 10^5 to 10^6 p.p.s., but with low throughput of light. Using a rotating mirror camera to traverse the image elements rates of 10^8 p.p.s. have been achieved for brilliant objects. Dissection by means of lenticular plates allows in many cases a considerably greater throughput of light. Simple cameras are now available that can take 300 pictures at 250 000 p.p.s. Lenticular plate image dissection has been applied to cine-micrography so that similar series of 300 pictures can easily be taken at 10^5 p.p.s. at magnifications up to 2000 ×. Alternatively, the use of fibre light guides for dissection allows long series of pictures of low resolution at 10^5 p.p.s. The combination of image dissection and deflecting image converter provides means for taking a series of 50 pictures at rates of the order of 10^9 p.p.s.

Using one or another of these techniques, reasonable photographs may be taken of most macroscopic phenomena that are not too remote, though synchronization problems are sometimes of overriding significance. The difficulties are greater when the interest is in fine spatial detail. These problems are accentuated in cine-micrography where the image moves many times faster than the event. In such cases one must take advantage of the most advanced techniques, and there is urgent need for the continued development of better methods.

§ 1. INTRODUCTION

A WIDE field of photography is customarily included in the term 'High-speed Photography'. It may be taken to include the recording of single-shot pictures with exposure times of less than one-thousandth of a second, and the taking of series of pictures at rates above 200 frames per second, and the taking of other records such as streak records which allow a time resolution shorter than a millisecond. The different branches of the subject might be classified by their historical or geographical development, or by their speed range, and so on ; but one of the most convenient classifications is by the kind of record that is made, and the method by which the recording is made.†

Many of the phenomena that need to be studied by high-speed photographic methods can be adequately 'frozen' by the use of a single short flash of light. That is, a single picture is sufficient. Many more require speeds of a few thousand frames per second. Compact and efficient cameras for this sort of job are commercially available. Such applications as these constitute the bulk of the work that is undertaken, to the extent even that they may be called the classical methods of high-speed photography. However, they are, just because of this, mentioned only briefly below. The emphasis of the presentation is on the newer techniques, their advantages, their limitations and their inherent possibilities. There is little discussion of the applications of the techniques, except in so far as is evident from the illustrations. A number of other volumes and reviews have been written presenting different aspects of the subject (Bourne 1948, Chesterman 1951, Collins 1957, Courtney-Pratt 1953 a, Fayolle and Naslin 1948, 1949, Jones 1952, Naslin and Vivie 1956, Schultze 1955, Society of Motion Picture and Television Engineers 1949–54, Spencer 1951, 1954).

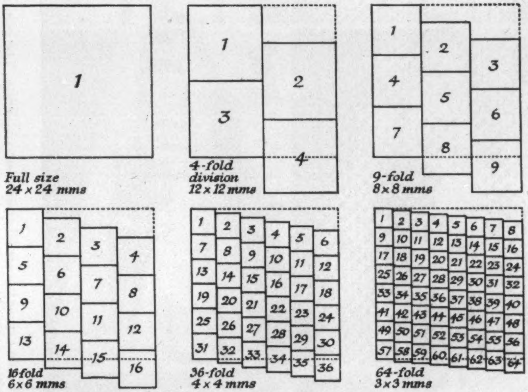
Before I begin to describe the various systems, I wish to draw attention to an important point. One can easily take pictures with a standard 35-mm cine-camera at 32 frames per second. These pictures are each 16 × 24 mm, and the quality of the pictures can be high. We may measure the quality as the number of black and white line pairs that can just be resolved across the frame in each dimension. One may expect to resolve in good 'still' cameras about 1000 line pairs across the frame each way. In commercial cinemas the resolution under ordinary projection conditions is rarely better than 350 line pairs from top to bottom of the picture on the screen. We still think of these as good pictures. However, if we drop by a further factor of two-thirds in resolution, the pictures are noticeably worse.

When we wish to increase the frame speed we are tempted to subdivide the picture area or reduce the film size, as shown in figure 1, as this would allow us to take more pictures per second without increasing the accelerations and the stresses in the film and/or the moving parts. If this sub-division is directly at the expense of the number of lines resolved across the frame, then little really has been gained. In any assessment of a technique of high-speed photography it is essential to consider both the framing rate and the resolution achieved.

---

† A brief survey of some of the methods of high-speed photography appears as part of an article I wrote some years ago, and I would like to thank the Editors of the *Journal of Photographic Science* for permission to make use again of matter that was published there.

Figure 2 shows six pictures taken by Professor Schardin (Schardin 1953). They are all of the same scene, but the resolution in each is different. The pictures have been arbitrarily divided up into blocks of constant density. The total number of picture elements is indicated on each picture. Where the number is only 500, i.e. about 20 lines × 25 lines, the picture quality is very low, and might well be taken as the lowest limit of useful picture quality. It would be just possible with a picture such as this to tell with an accuracy to 4 or 5% the position of large and obvious items in the field of view. On the other hand, where the number of picture elements is 20 000, i.e. about 125 × 160, the quality, subjectively judged, is quite good. We have usually taken, as a guide, that we should aim at a minimum resolution of 200 × 200 for high-speed pictures, though if severely pressed a figure of 70 × 70 is still acceptable, as may be seen by reference to the fifth of Schardin's pictures, which has only 5000 picture elements.



Note: The figures indicate picture sequence order

Figure 1. A diagram illustrating the principle of frame division. It is possible to achieve a larger number of small frames per second without increasing the accelerations and stresses in a camera mechanism. The advantages are however slight if this is achieved at the expense of proportionate loss of picture quality. (*Reproduced by courtesy of* W. Deryck Chesterman.)

Still higher picture quality is almost always worth aiming for, particularly where one is able to make the subsequent analysis or measurement directly on the primary negative. I have added this proviso, as only too frequently the quality of a really high resolution picture is degraded by inadequate playback, measuring, printing or copying equipment.

In the sections that follow there will be frequent reference to spatial resolution and to time resolution. The choice of a suitable means for studying any event depends primarily on these two resolutions; but there are other limitations, and the most serious of these are available light level at the photographic emulsion,

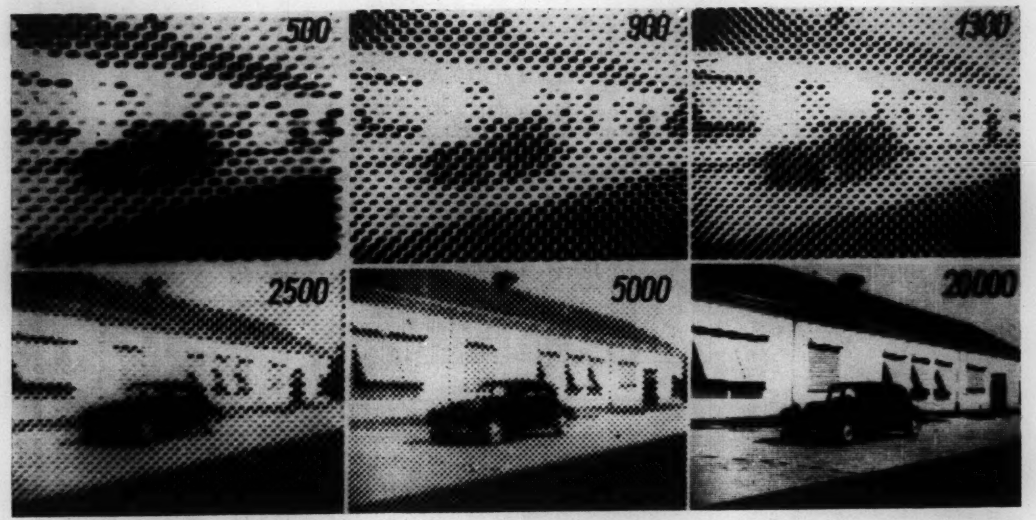


Figure 2. These six pictures are all of the same scene but are all of different resolution. The pictures have been divided up into blocks of constant density. The total number of these blocks in each picture is indicated. (*Reproduced by courtesy of* Professor H. Schardin.)

synchronization difficulties, length of sequence required, repeatability of the event, convenience or otherwise of operation, labour in reading the recording, and cost.

No one of the following methods, nor even any single combination of them, can provide a universal answer. It is often the best plan to make a qualitative overall survey by some simple means, and to study the particular feature of greatest interest by some more specialized but limited technique.

§ 2. STREAK RECORDS

For streak records, the attention is confined to events that occur along some chosen line. The image of this line is focused on the sensitive surface of photographic film which is moved during the event. The result is thus a record of the development of the phenomenon along the chosen line as a function of time. A simple diagram is shown in figure 3. Continuous film cameras have writing speeds of not more than a few tens of metres per second, and the minimum time between two events that may just be resolved is 10^{-5} sec (Chesterman 1951, Chesterman and Myers 1951, Hercock 1947, Jones 1952). In a recent camera a short loop of film is accelerated and, as its inertia is so much less, time delays and film wastage are much reduced (Prudence 1957). With drum cameras the writing speed can be 100 or 200 m/sec, and the time resolution 10^{-6} sec or a little less (Chesterman 1951, Courtney-Pratt 1952, Courtney-Pratt and Thackeray

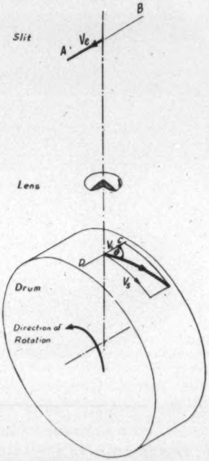


Figure 3. A diagram of a simple drum camera. From measurements of the streak record, velocities of movement of the object can be determined relative to the speed of movement of the film. (*Reproduced by courtesy of the ' Photographic Journal '.*)

1956 b, Eyles 1941, Dr. Frank Früngel G.m.b.H. (C.L.)†, Hercock 1947, Jones 1928, Jones 1952, Payman, Shepherd and Woodhead 1937, Weibull 1947).

Instead of a fixed image and moving film, the beam of light may be reflected from a rotating mirror or prism so that the line image moves over film that is held stationary. A writing speed of a thousand metres per second is easily possible with a lens aperture of f/8, and higher speeds have been attained at the expense of smaller light-gathering power (Adams 1950, Armament Research Establishment 1952 b, Beckman and Whitley Inc. (C.L.), Brixner 1957, Cairns 1944, Chesterman 1951, Croney 1948, Herzberg and Walker 1948, Jones 1952, Payman *et al.* 1937). In fact the ratio of writing speed to numerical aperture is a convenient figure of merit and a value of 150 m/sec has rarely been exceeded, for any drum camera, or a value of 1000 m/sec for any mirror camera.

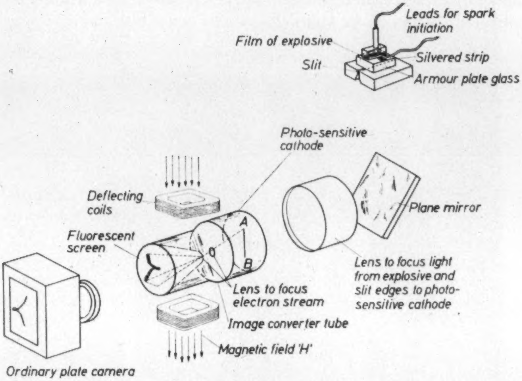

Leads for spark initiation
Film of explosive
Slit
Silvered strip
Armour plate glass
Photo-sensitive cathode
Deflecting coils
Plane mirror
Fluorescent screen
A
B
Lens to focus light from explosive and slit edges to photo-sensitive cathode
Lens to focus electron stream
Image converter tube
Magnetic field 'H'
Ordinary plate camera

Figure 4. A diagram of image converter tube apparatus used to obtain streak records of explosive reactions. (*Reproduced by courtesy of ' Research '.*)

Professor Schardin at the recent Congress on High-speed Photography (Schardin 1957 b) considered the ultimate limits to which one might press a rotating mirror streak camera. The minimum width of an image of the slit is determined by diffraction and the aperture of the system. Using, for the mirror, materials of the highest specific strength available, the peripheral speed of the mirror can hardly be greater than 500 metres/sec. From this it follows that the time resolution cannot be better than $\frac{1}{4} \times 10^{-9}$ sec, whether one uses a high numerical aperture and correspondingly low writing speed, or any other combination. Mirror streak cameras built so far, fall considerably short of this ultimate limit.

It is possible to avoid all mechanically moving parts. If the image of the chosen line is focused on the photosensitive surface of an image converter tube a corresponding line image would appear on the fluorescent screen. The electron stream can be deflected by a changing transverse electric or magnetic field to provide a fast time-base. The record on the fluorescent screen of the whole event can be photographed by an ordinary stationary camera (Courtney-Pratt 1949, 1950, 1951, 1952, Courtney-Pratt and Thackeray 1956 b, Jenkins and Chippendale 1951, Richards 1952). The scheme is shown diagrammatically in figure 4, and a streak record of the explosion of some cadmium azide in figure 5(*a*). This may be compared with figure 5(*b*) which is an example of a record taken with a very fast rotating mirror camera. Figure 6 is a view of a versatile commercially produced image converter camera that can take streak records, single-shot or repetitive short exposures. In Cambridge we have attained writing speeds of 300 000 m/sec with an effective aperture of f/4, so that the figure of merit is some 75 times greater than for the best mechanical cameras (Courtney-Pratt 1952), and already the minimum time resolution is shorter than the ultimate limit for rotating mirror cameras calculated by Professor Schardin. Recently some Russian workers (Zavoiskii and Fanchenko 1955) have made use of these ideas

---

† References to commercial literature are indicated by (C.L.).

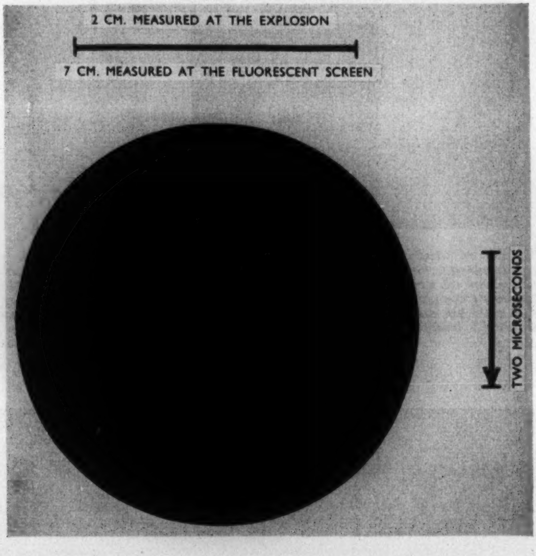

2 CM. MEASURED AT THE EXPLOSION
7 CM. MEASURED AT THE FLUORESCENT SCREEN
TWO MICROSECONDS

Figure 5(*a*). A streak record showing the detonation of powdered cadmium azide. This picture was obtained at our laboratory in Cambridge, using an image converter tube and magnetic time-base. The writing speed measured at the fluorescent screen was 17 000 m/sec, the width of the image of the slit at the screen was ⅓ mm, and the lateral magnification was 3·5 ×. (*Reproduced by courtesy of* G. T. Rogers.)

and increased the writing speed by a factor of several hundred and the time resolution correspondingly.

Streak records, though they give less total picture detail, do usually allow a finer analysis of time detail than other forms of photography.

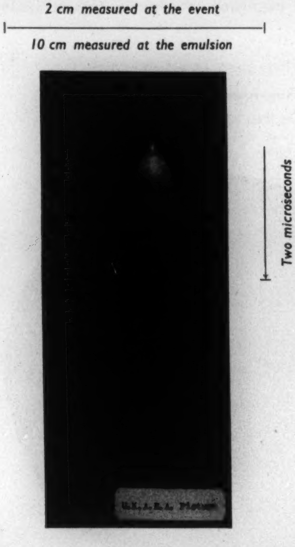


Figure 5(*b*). U.K.A.E.A. picture. Figure 5(*b*) may be compared with figure 5(*a*). It is an example of a streak record of a fast self-luminous phenomenon taken on a rotating mirror camera. Writing speed at the emulsion was 26 000 m/sec. The width of the image of the slit at the emulsion was 0·5 mm. The optical magnification from object to image was 5 × and the record is reproduced here half-size. (*Reproduced by courtesy of* E. W. Walker and U.K.A.E.A.)

§ 3. SINGLE EXPOSURES

The single-exposure record is an ordinary two-dimensional photograph. Mechanical shutters are not practicable for exposure times less than about a millisecond, though some advance is possible at the expense of aperture.

Electronic flash tubes are available that give a high enough light output for photography by reflected light and an effective exposure as short as two or three microseconds (Aldington and Meadowcroft 1948, Bourne 1948, Chesterman 1951, Chesterman and Glegg 1956, Edgerton 1946, Edgerton and Killian 1939/54, Jones 1952, Laporte 1956, Lennard-Jones and Mitchell, patent, Marshak 1957). A stereoscopic pair of photographs taken with flash illumination is shown in figure 7. A simple spark discharge can be made to give as much light and as

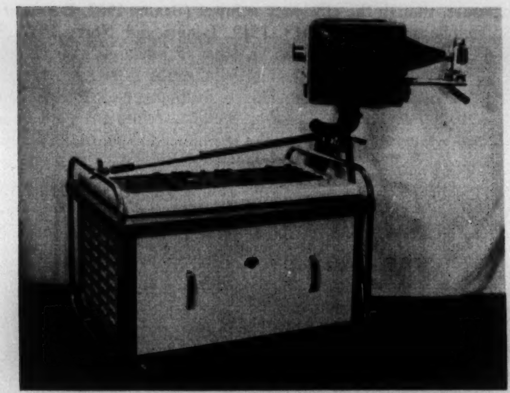


Figure 6. A general view of a complete image converter camera and control console. This equipment can take streak records at writing speeds up to 150 000 m/sec, two-dimensional pictures with exposure time from 10^{-6} sec to 1 sec, or repetitive exposures for stroboscopic viewing at rates up to 10^{5}/sec. Complete time delay and synchronization circuits are available. More recently, single-purpose equipment of much improved performance has been developed and is also available commercially. (*Reproduced by courtesy of* Mullard Electronic Research Laboratories Limited.)

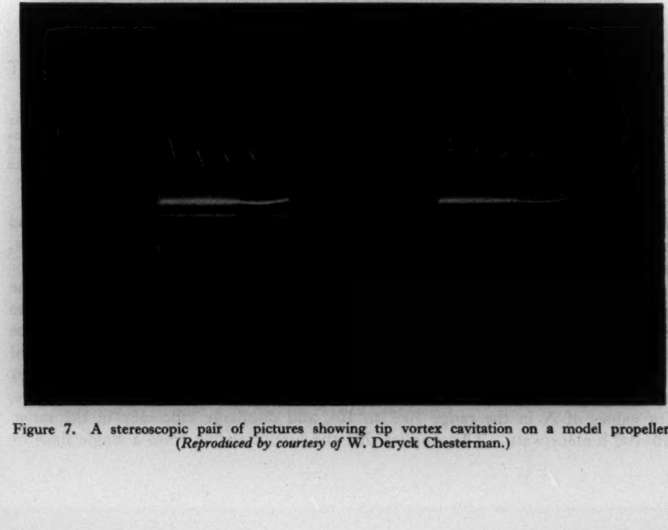


Figure 7. A stereoscopic pair of pictures showing tip vortex cavitation on a model propeller. (*Reproduced by courtesy of* W. Deryck Chesterman.)

brief an exposure, though the efficiency is lower (Beams 1937, Beams *et al.* 1947, Boys 1893, Fayolle and Naslin 1948, 1949, Looms and North 1957, Standring and Looms 1952, Whitehead and Fance 1945). If, however, a less intense source is inadequate, as for transparence photography, much briefer sparks are possible (Fayolle and Naslin 1948, 1949, Fitzpatrick, Hubbard and Thaler 1950, Quinn and Bourque 1951). A spark shadowgraph of a bullet in flight is shown in figure 8. Three wind-tunnel photographs reproduced as figures 9, 10 and 11 also illustrate the detail that can be achieved by microsecond spark exposures, and show the different features that can be accentuated by the use of interferometer, schlieren and shadowgraph techniques (Holder and North 1956).

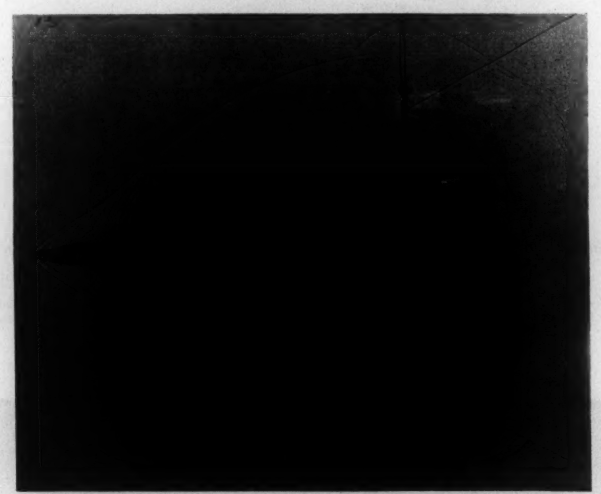


Figure 8. A spark shadowgraph of a bullet that has penetrated a brittle sheet. This shows the great detail and clarity that can be achieved. This is one of the best and simplest methods provided a silhouette photograph is sufficient. The shock waves and the turbulence caused by the bullet and the fragments are clearly visible. (*Reproduced by courtesy of the* Laboratoire Central de l'Armement, Paris.)

The luminous efficiency of air sparks can be increased by increasing the pressure or the path length so that the effective impedance is higher. This has been done by a number of workers by guiding the breakdown along the surface of an insulator (Aumont and Vodar 1956, Fayolle and Naslin 1948, 1949, Tawil 1957). Very roughly, if the energy to be dissipated in the discharge tube or spark is N joules the effective exposure time can be kept as small as $N/10^7$ seconds, for values of N in the range 10^{-1} to 10^4, and discharge efficiency in the range 10–100 lumens/watt. This implies that the luminous flux from a single flash or

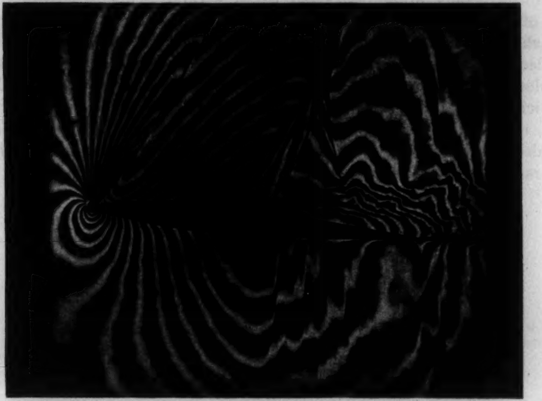


Figure 9. Interferometer photograph of the flow past a two-dimensional aerofoil. Exposure time one microsecond. (*Reproduced by courtesy of the* Director of the National Physical Laboratory.)

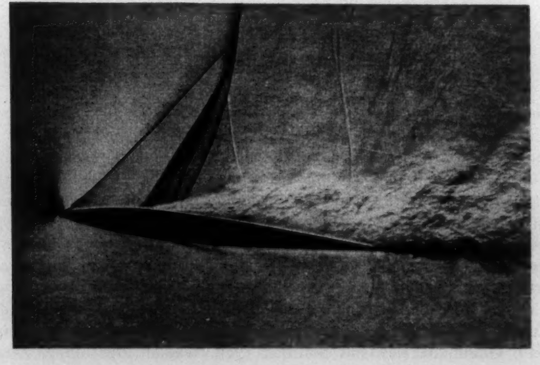


Figure 10. Toepler schlieren photograph of the flow past a two-dimensional aerofoil. The direction of the knife edge was horizontal. Exposure time one microsecond. (*Reproduced by courtesy of the* Director of the National Physical Laboratory.)

spark source, of the kind so far used as illuminants, is rarely if ever in excess of 10^9 lumens. This would not seem to be a fundamental limitation, and this aspect of the subject would warrant further research.

Flashes as brief as a quarter of a microsecond and of higher total intensity than electrical discharges can be produced by the passage of a shock wave from an explosion through a gas such as argon (Grimshaw and Hardy 1951, Muraour and Michel-Lévy 1934, 1949, Winning and Edgerton 1952). Direct, pyrotechnic flashes, though they can be vastly brighter, are not nearly as short. Fine wires 'exploded' by the discharge of a condenser can provide intense sources for times in the range 10 to 100 microseconds and longer (Conn 1951, Thackeray 1957).

Figure 11. Direct-shadow photograph of the flow past a two-dimensional aerofoil. Exposure time one microsecond. (*Reproduced by courtesy of the* Director of the National Physical Laboratory.)

No form of flash illumination can be used to control the exposure if there is a high level of background lighting or if the object is strongly self-luminous. Electronic shutters are free from this limitation. A Kerr cell shutter is shown diagrammatically in figure 12, and an actual equipment in figure 13. Such a shutter can be operated in a tenth of a microsecond, and figure 14 is a typical picture of a bullet in flight. The energy to operate such a cell increases with the aperture; it is progressively more difficult to operate cells of larger aperture in such a short time, though smaller cells can be operated faster (Cady and Zarem 1948, Funfer and Muller 1956, Heine-Geldern 1956, Karolus 1956, Prime and Saxe 1949, Quinn, McKay and Bourque 1950, Walker 1956, 1957, Zarem and Marshall 1950, Zarem, Marshall and Poole 1949). Magneto-optic shutters have advantages for exposures of a microsecond and longer (Edgerton and Wyckoff 1951). An image converter tube can be used as a shutter if the accelerating potential is applied as a pulse of short duration (Chippendale 1952, 1957, Courtney-Pratt 1949, 1950, 1951, 1952, Courtney-Pratt and Thackeray 1956 b, Gibson, Bowser, Ramaley and Scott 1954, Jenkins and Chippendale 1951, Meek and Turnock 1952, Richards 1952, Saxe 1957, Saxe and Chippendale 1955, Turnock 1951). Figure 15 is a one-microsecond exposure of a bullet in flight and may be compared with figure 14. Pictures have been obtained with a resolution of 600 lines across the field and an exposure time of 1/20 microsecond (Jenkins and Chippendale 1951). The sort of resolution attainable is shown in figure 16. Recently image converter tubes have been used by Chippendale and Saxe to take pictures with an exposure time of about 3×10^{-9} sec (Saxe 1957, Saxe and

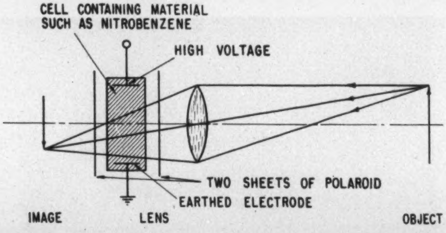


Figure 12. A simplified diagram of a Kerr cell for use as a photographic shutter. In the closed position the sheets of polaroid are so oriented that no light passes. A change of electric field affects the plane of polarization, and some light can then pass through the shutter.

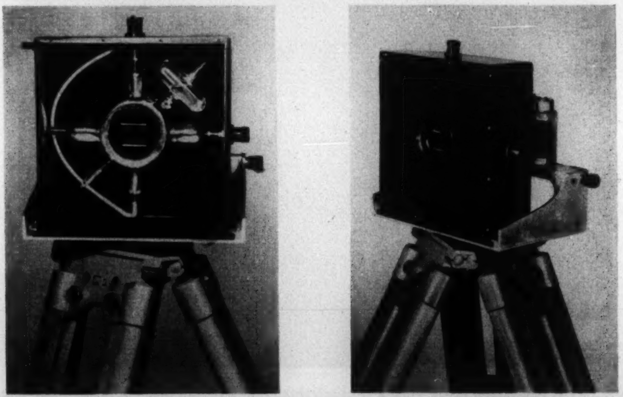


Figure 13. Two views of a simple Kerr cell shutter for use with a 35-mm camera. The clear aperture is 15 × 15 mm. It can give exposures as short as 10^{-7} sec. (*Reproduced by courtesy of the* Laboratoire Central de l'Armement, Paris, *and of 'J. Soc. Mot. Pict. Telev. Engrs'.*) (Fayolle and Naslin 1954.)

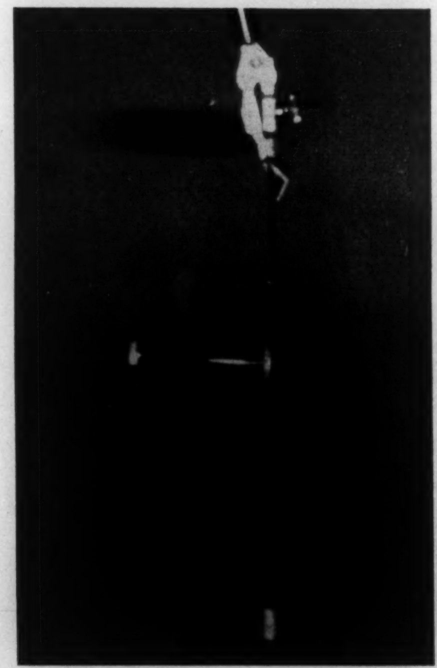


Figure 14. A photograph taken, using a Kerr cell shutter, of a 20 mm bullet in flight. The shutter was phased to take advantage of the peak light output of a ' Defatron ' guided spark discharge. (*Reproduced by courtesy of the* Laboratoire Central de l'Armement, Paris, *and of* '*J. Soc. Mot. Pict. Telev. Engrs*'.) (Fayolle and Naslin 1954.)

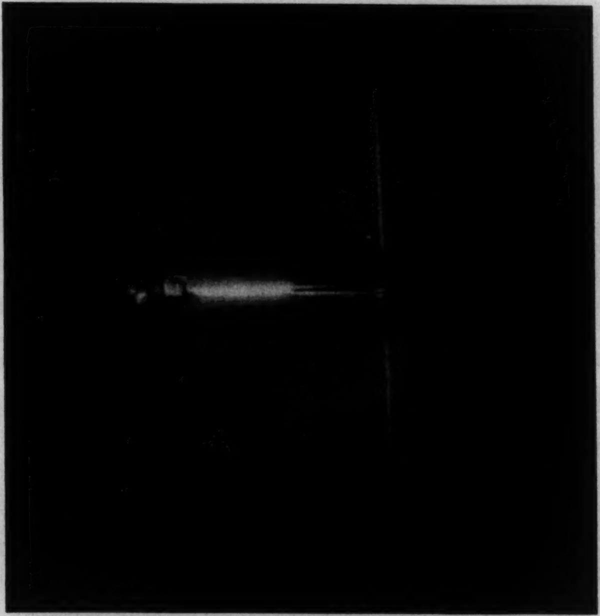


Figure 15. This was taken using an image converter tube as a shutter, with an exposure time of one microsecond. The image converter shutter was phased to operate near the peak of an 'Arditron' flash. The bullet was 20 mm diameter, and was travelling at 2400 ft/sec. A number of grooves milled on the shoulder of the round to provide photographic detail are clearly visible. (*Reproduced by courtesy of* D. P. C. Thackeray and J. Brickell.)

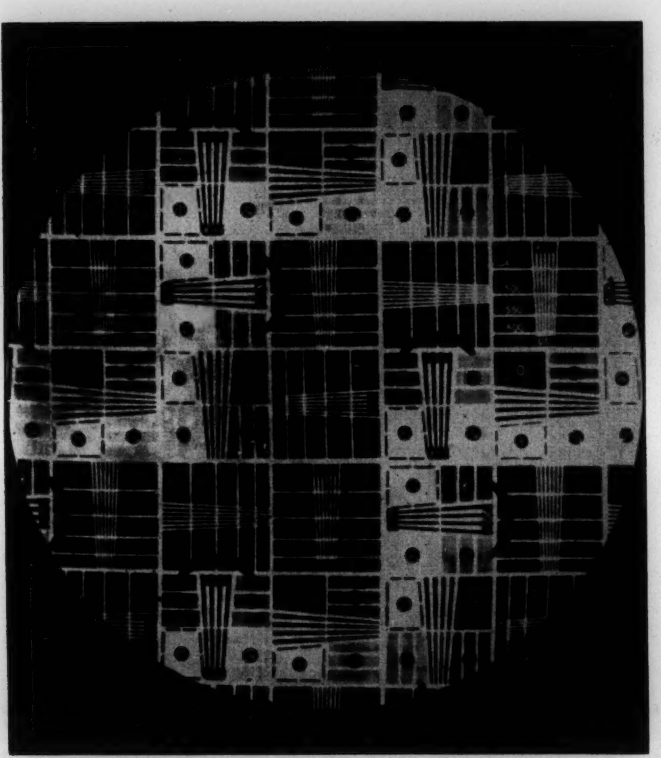


Figure 16. A photograph of a resolution chart through a standard Mullard image converter tube. It is possible to achieve a resolution of 30 line pairs per millimetre all across the fluorescent screen, which is 10 cm in diameter. With the wide-aperture lenses that one almost always uses in high-speed photography the resolution is severely restricted below this by the lenses, and one rarely achieves more than the 500 line pairs which are clearly resolved in the original of this illustration. (*Reproduced by courtesy of* R. A. Chippendale and J. A. Jenkins, *and* Mullard Electronic Research Laboratories Limited.)

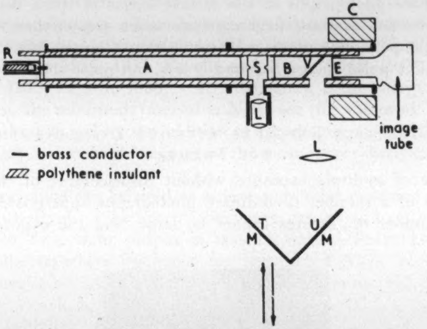


Figure 17. Experimental arrangement of apparatus used for photographing the earliest stages of an electric discharge using an image converter tube to give an exposure time of the order of 10^{-9} sec. The phase of the spark relative to the exposure may be delayed by moving the mirrors MM so that the light from the spark must take a longer path and arrive later. (*Reproduced by courtesy of* R. F. Saxe and R. A. Chippendale, *and* '*Brit. J. Appl. Phys.*')

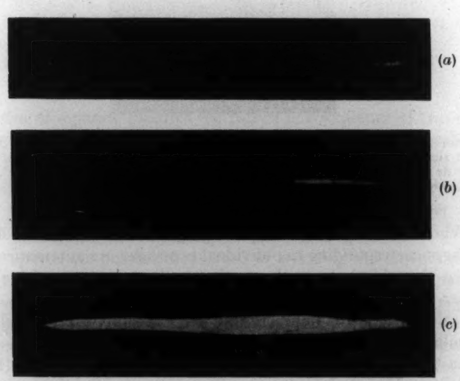


Figure 18. Three exposures, taken with the image converter equipment shown in the preceding figure, of the earliest stages of an electric spark. The exposure time in each case was 3×10^{-9} sec, and the delay between commencement of breakdown and the mid-point of the exposure is in the three cases: (*a*) 0, (*b*) $1{\cdot}35 \times 10^{-9}$ sec, (*c*) $3{\cdot}4 \times 10^{-9}$ sec. (*Reproduced by courtesy of* R. F. Saxe and R. A. Chippendale, *and* '*Brit. J. Appl. Phys.*')

Chippendale 1955). The layout of their apparatus is shown in figure 17, and in figure 18 three photographs of the earliest stages of spark discharges. The effective aperture and available field angle are much greater than for a Kerr cell. There is even the possibility of image intensification (Chippendale 1957, Courtney-Pratt 1952). The pulse energy required is less, but pulse shape is important.

§ 4. MULTIPLE EXPOSURE WITHOUT DISPLACEMENT OF THE FRAMES

In the case of multiple exposure without displacement of the frames, the record consists of a number of ordinary photographs superposed on the same plate. The number of pictures cannot be large, and the exposure latitude is reduced. Where the object to be photographed is moving rapidly across the field, the images corresponding to individual exposures are automatically displaced. To the extent that this reduces the confusion due to superposition of frames, a larger number of exposures may be recorded.

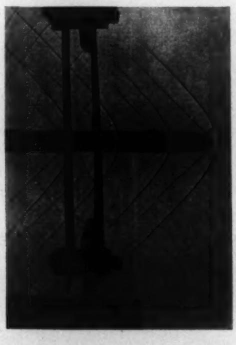


Figure 19. Four spark shadowgraphs all taken on the one plate by means of four discharges passed in rapid succession through a single spark gap. (*Reproduced by courtesy of the* Laboratoire Central de l'Armement, Paris, *and of* '*J. Soc. Mot. Pict. Telev. Engrs*'.) (Fayolle and Naslin 1954.)

Mechanical shutters are impracticable for exposure times much less than a tenth of a millisecond. A series of flashes or sparks can be used. Much of the finest work in this field has been done by Professor H. E. Edgerton (Edgerton and Killian 1939/54). If entirely separate sources are provided these may be as brief and as intense as for single exposures and may be fired at any chosen intervals. Equally, with separate controls separate storage condensers may be arranged to discharge through a single gap (Adams 1951, Armament Research Establishment 1952 a, Luy and Schade 1956, Fayolle and Naslin 1954). Four such superposed photographs are shown in figure 19.

If simple discharges are to be produced repetitively between the same electrodes the minimum time between exposures is limited by the de-ionization characteristics of the surrounding gas. With high-energy sources the time between successive discharges must be longer (Adams 1951, Chesterman 1951, Chesterman and Peck 1952, Edgerton 1952, 1957, Fayolle and Naslin 1948, 1949, Früngel 1956, Germeshausen 1949, Laporte 1956). Approximately, if the energy dissipated per flash is N joules, the minimum interval is $N/10^4$ sec (for values of N in the range 10^{-1} to 10^4). Dr. Früngel by the addition of a hydrogen quenching gap in series with the spark gap has managed to increase the mean power dissipation several fold (Früngel 1957).

A cathode-ray tube can be used as a flash source for exposure times down to about one microsecond. The specific brilliance can be as high as for a carbon arc, though the total light output is usually small. There are considerable advantages as the repetitive frequency can approach 1 Mc/s, and synchronizing delays can be negligible (Amiot (C.L.), Brookes and Monro 1957, Courtney-Pratt and Thackeray 1956 a, b, Devaux 1956, Ferranti Ltd. (C.L.), General Electric Co. Ltd. (C.L.), Philips', N.V. (C.L.)).

Electronic shutters can be operated much faster than repetitive spark illumination (Walker 1956, 1957, Froome 1948, 1952). For example a Kerr cell shutter of aperture 0·1 cm^2 has been operated with an exposure time of 10^{-7} sec for ten or twenty exposures at intervals of 0·4 microsecond (Froome 1948, 1952). The repetitive Kerr cell equipment described in § 5 (*e*) below could equally well be used for superimposed pictures. Image converter tubes of high aperture have been operated with exposure times of a microsecond at a frequency of 100 000/sec for long periods (Richards 1952). For the same-sized driving gear they can be made to operate faster than Kerr cell shutters.

§ 5. SERIES OF SEPARATE PICTURES

For the series of separate pictures the record is a sequence of two-dimensional pictures. There are a number of general methods, and considerable diversity with each.

(*a*) The best known is ordinary cinematography—intermittent movement of the film and a mechanical shutter. Frame speeds up to 300 per second and exposure times down to a tenth of a millisecond are possible.

(*b*) Higher speeds are possible if, instead of arresting the film for each picture, the exposure time is made so short that movement within the exposure time is negligible. For a resolution of 200 lines across the frame, the exposure time must not be greater than 1/200 of the time between frames. Using repetitive spark illumination it is thus possible to take transparence photographs at 10 000/sec. Photographs by reflected light with the higher light levels required are difficult even for close phenomena at more than 5000/sec (Chesterman 1951, Fayolle and Naslin 1948, 1949, Früngel 1956, Henry 1944, Jones 1952, Luy and Schade 1956).

Dr. Früngel, using his hydrogen quenching gap in series with the gap that produces light, can produce a burst of flashes at 25 000/sec with an energy per flash of the order of one joule (and higher energies are possible at lower speeds).

With this equipment and a straightforward drum camera shown in figure 20 he has been able to take good pictures, even by reflected light, at rates up to 8000/sec, and pictures of reduced frame height at rates up to 25 000/sec (Früngel 1957). An indication of the picture quality achieved can be obtained from figure 21. Instead of using a drum or mirror camera, film can be wound continuously from a magazine. For 16 mm × 25 mm frames the maximum speed is then about 2500/sec, and a typical example is shown in figure 22. Again, a cathode-ray tube can sometimes be a convenient source (Devaux 1956).

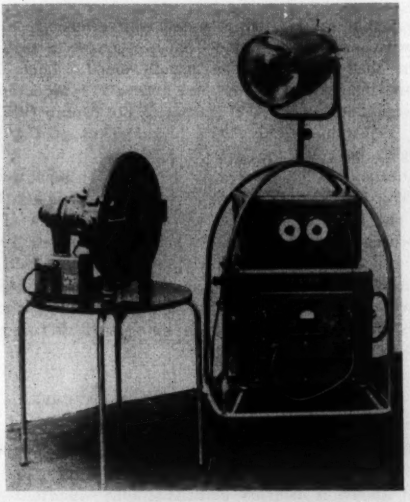


Figure 20. A drum camera and the ultra-rapid flasher designed by Dr. Früngel. The drum is capable of a linear film speed of 80 m/sec ; and the repetitive flash unit can produce 1-joule flashes at 25 000 flashes per second. Higher energy flashes are possible at lower speeds. (*Reproduced by courtesy of* Dr. Ing. Frank Früngel, G.m.b.H.)

Electronic shutters could be used instead of intermittent illumination with some advance in maximum speed of operation, and even for objects that are self-luminous, or where there is strong ambient lighting. Where only a few frames are required it is possible to use an image converter tube both as the shutter and, by applying suitably phased deflection fields, as the means of separating the frames (Courtney-Pratt 1949, Sultanoff 1956). The fluorescent screen can then be photographed by an ordinary stationary camera. As there is no movement of any individual picture on the fluorescent screen relative to the photographic emulsion it does not matter whether the phosphor has a fast or a slow decay time. On the other hand, when there is continuous movement of the film and an image converter tube is used simply as a repetitive shutter, it is important that the decay time be short.

(*c*) The stringency of the requirements of shutter operation or brevity of light flash can be relaxed without necessitating arrest of the film for each exposure if some form of compensation is introduced so that during the exposure the image is moved forward at the same speed as the film. The compensation is usually effected by reflection of the beam from a rotating mirror (Scophony Ltd. 1946, Suhara 1934, Uyemura 1957), or by refraction as it passes through a rotating parallel-sided block or prism (Boon 1944, Chesterman and Myers 1951, Eastman Kodak (C.L.), Herden 1957, Smith 1945, Waddell 1946, 1954, 1955, 1957, Western Electric (C.L.), Wollensak (C.L.)) as shown in figure 23, or by use of a rotating ring of lenses (Geary 1939, Merlin, Gerin, Dubuit (C.L.), Thun 1956, Vinten Ltd. (C.L.)). With full 35 mm frame size (16 × 25 mm) the repetition rate can be 2000/sec (Chesterman and Myers 1951) or 2500/sec (Waddell 1954). With

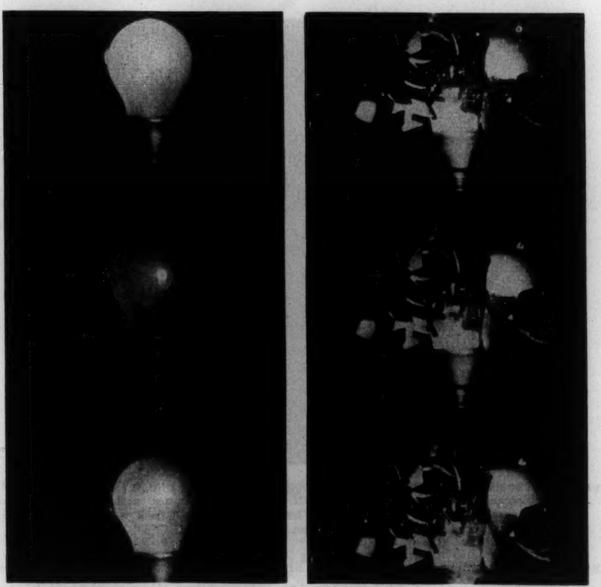


Figure 21. Six pictures enlarged about 4× from a series taken at 8000 frames per second on a simple drum camera without image compensation, using a repetitive spark source. The camera and source are those shown in figure 20. The effective exposure is obviously less than 10^{-6} sec. (*Reproduced by courtesy of* Dr. Ing. Frank Früngel, G.m.b.H.)

reduced frame size the repetition rate may be correspondingly increased, so that using 16-mm film and pictures of half the usual frame height (i.e. 3·7 mm), a speed of 20 000 pictures per second is commonly achieved (Waddell 1955, Hadland 1957). The exposure time may be short or may be increased to about half the time between successive frames. The long exposures, as in all forms of photography, increase the uncertainty of the measurement of the position of moving objects, but make it much easier to provide adequate illumination. Because of this and because image compensation is always subject to some error, some cameras incorporate compensation and means for intermittent illumination (Chesterman and Myers 1951). Figure 24 shows a sequence of pictures taken in this way with the Acmade camera and synchronized flash unit.

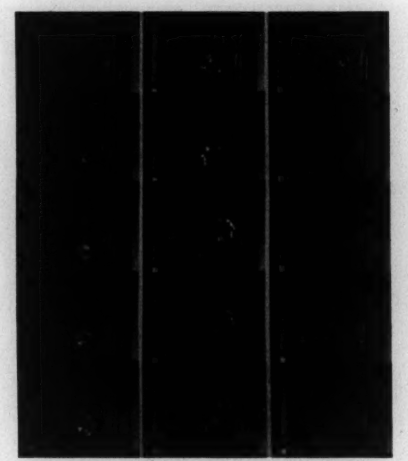


Figure 22. A sequence of pictures taken on continuously moving film at 1000 per second showing the coalescence, growth and collapse of small transient cavities in water. Illumination was provided by a repetitive flash source. The duration of each flash was 5×10^{-6} sec. (*Reproduced by courtesy of* W. Deryck Chesterman.)

A number of cameras can be used at the same time, and by displacing the phase of their shutter operation greater numbers of pictures per second can be obtained (Wilkinson and Romig 1954). In usual practice, though separate objectives are required, considerable economy is possible by arranging that a large part of the apparatus is common to all the optical trains (Baird 1946, Baird and Durie 1949, Merlin, Gerin, Dubuit (C.L.)).

A different method of image compensation has been developed by C. D. Miller (Miller and Shaftan 1952, Miller and Scharf 1954). The principle is shown in figure 25. A primary image is formed on a corner reflector. Translation of the reflector displaces the reflected beam and introduces displacement of a secondary image which can conveniently be matched to the movement of the film by mounting the film and the array of reflectors on a single rotating drum. These cameras are simple to operate and can take about 500 pictures, 4 mm × 5 mm, on a length of conventional 8-mm cine-film at rates up to 100 000/sec, with a resolution of 20 lines/mm (i.e. a resolution of 80 × 100 lines across the field).

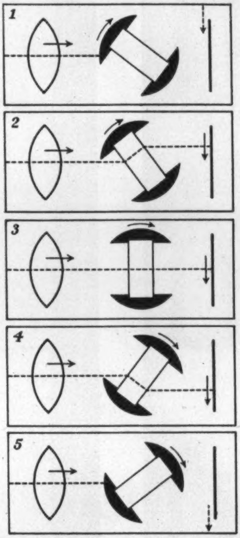


Figure 23. A diagram showing the principle of image compensation by means of a rotating block of glass.

(*d*) Even higher speeds are possible if effectively separate cameras are used for each frame. There need then be no movement of the film. The Marley camera (Jones and Eyles 1949), shown diagrammatically in figure 26, is typical of one group of such designs. The lens of each camera focuses directly on the object, and the cameras are successively exposed by the movement of slots in a rotating disc. About 60 frames can be recorded at 60 000/sec, though the effective numerical aperture is low.

(*e*) In another group of designs that have separate cameras for each frame a main lens forms a real image of the object on a rotating mirror which directs the light successively to the separate cameras (Armament Research Establishment 1951, Beckman and Whitley Inc. (C.L.), Brixner 1954, 1956, Jacobs and Klebba 1950, Miller 1949, Scophony Ltd. 1947, Stanton and Blatt 1948, Sultanoff 1956,

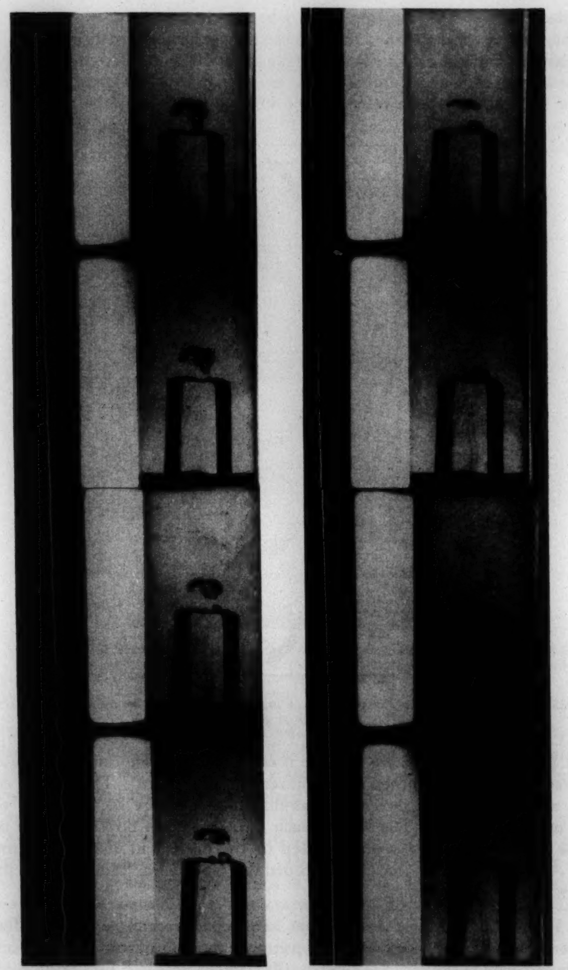


Figure 24. An 8-frame sequence showing the collapse of a steam bubble cavity in water. This series was taken with the Acmade camera at 1250 p.p.s. using synchronized flash illumination so that each exposure is of 5 microseconds duration. The oscillograph trace on the film indicates pressure changes on a submerged hydrophone. The picture frames lag the pressure trace by 6 frames. (*Reproduced by courtesy of* W. Deryck Chesterman and D. R. Glegg.)

Tchernyi 1957, Walker 1956, 1957, Zernow and Hauver 1957). The scheme is shown in figure 27. In the earliest apparatus working on this principle four optical trains are used in multiplex (Miller 1949). One of these cameras can

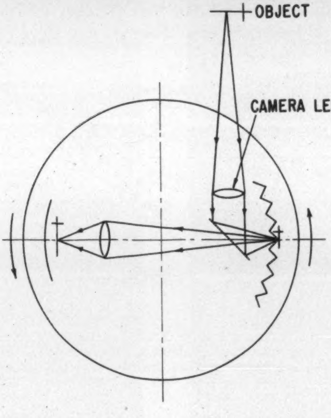

OBJECT
CAMERA LENS

Figure 25. Diagram showing the principle of operation of the first of the designs by C. D. Miller. The primary image is formed on the row of reflector pairs which are rotating with the drum. The reflected primary image moves at twice the speed of the drum. The secondary image, of half the size of the primary image and moving at half the velocity of the reflected primary image, falls on film which also is mounted on the drum and moves with the film. (Miller and Shaftan 1952, Miller and Scharf 1954.)

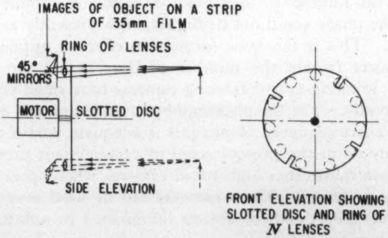

IMAGES OF OBJECT ON A STRIP OF 35mm FILM
RING OF LENSES
45° MIRRORS
MOTOR
SLOTTED DISC
SIDE ELEVATION
FRONT ELEVATION SHOWING SLOTTED DISC AND RING OF N LENSES

Figure 26. A diagram showing the principle of the Marley camera. With one slot in the disc all the lenses would be exposed once in one revolution. With R slots in the disc the N lenses would all be exposed in $1/R$ of a revolution, and provided N and R have no common factor. Simultaneous exposures do not occur.

take 170 frames 1 cm × 1·4 cm with a lens aperture of f/26 at 3 000 000/sec (Brixner 1954); a newer model can take 96 pictures at $1{\cdot}5 \times 10^7$/sec (Brixner 1956). For other cameras working on this principle the exposure time has been controlled by adding a repetitive Kerr cell shutter (Walker 1956, 1957, Armament Research Establishment 1951). In one of these, eighty 50-mm diameter frames may be taken at 500 000/sec with exposure times of 10^{-7} sec though the numerical aperture is less than f/50 (Walker 1956). Another will take 30 pictures, each 5 mm × 8 mm, at 10^7/sec with exposure times of 2×10^{-8} sec (Walker 1957). A number of other cameras of this general type have been designed and some excellent equipment is commercially available from Beckman and Whitley Inc., San Carlos, California. An example (Zernow and Hauver 1957) of work possible with one of these cameras is shown in figure 28.

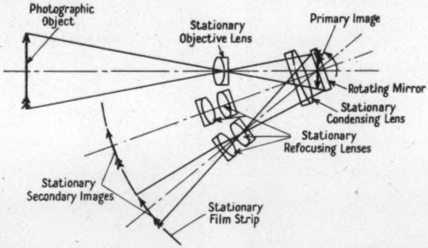

Photographic Object
Stationary Objective Lens
Primary Image
Rotating Mirror
Stationary Condensing Lens
Stationary Refocusing Lenses
Stationary Secondary Images
Stationary Film Strip

Figure 27. Schematic diagram of the principle of cameras designed independently by I. S. Bowen and C. D. Miller. The image of the object is formed on the surface of a rotating mirror. The reflected beam of light is successively directed toward a number of camera lenses, each of which records one image of the object on a stationary strip of film. (Miller 1949, Stanton and Blatt 1948.)

Professor Schardin (Schardin 1957 b) considered the fundamental limitations of these rotating mirror framing cameras and showed that the product of the framing rate and the number of lines resolved across the frame in the direction of movement of the image could not be higher, with presently available materials, than 4×10^9 sec^{-1}. This is the same formula in effect as applies to the rotating mirror streak camera (where the number of lines across the frame has been reduced to one). Rotating mirror framing cameras have small effective numerical aperture and are costly. For the photography of self-luminous events, if expense is no object, if a short sequence of pictures is adequate, and if there is plenty of light to spare, they do at the present stage of development produce more satisfactory pictures than most other high-speed cameras when operating in the range 10^5–10^7 pictures per second. These cameras can be used over a wide range of magnifications ($1/\infty$–$100/1$) and for objects illuminated in reflection or silhouette as well as for self-luminous objects.

(*f*) It is quite possible to assemble a number of separate cameras each fitted with its own electronic shutter, and to take a series of pictures by operating the shutters one after the other. The limitations for each on exposure time and aperture are then as described in § 3. The throughput of light and the resolution can be as good at least as for rotating mirror framing cameras. The interval between frames can be briefer. The pictures necessarily are each taken from a

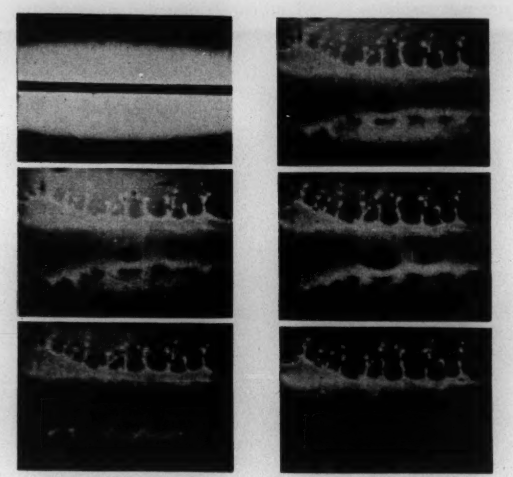


Figure 28. An example of cinemicrography with a Beckman and Whitley high-speed framing camera. These are selected frames from a sequence of an exploding tungsten wire. A static picture of the wire and frames 16–20 are included. Magnification 16 × ; framing rate $1{\cdot}2\times10^6$ frames/sec. A 4 μF condenser at 2 kv was discharged through the wire, which was ½ in. long and 0·002 in. in diameter. In the pictures one can see the development of numbers of fine sprays of small droplets of molten material. (*Reproduced by courtesy of* Aberdeen Proving Ground, M. Sultanoff, and L. Zernow and G. E. Hauver.) (Zernow and Hauver 1957.)

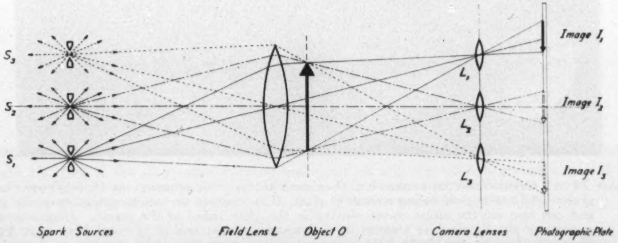

S_3
S_2
S_1
Image I_1
Image I_2
Image I_3
L_1
L_2
L_3
Spark Sources
Field Lens L
Object O
Camera Lenses
Photographic Plate

Figure 29. A diagram showing the principle of the Cranz–Schardin system. The camera lens L_1 records an image of the object when source S_1 is fired.

Figure 30. A Cranz–Schardin system has been used to take this sequence of 18 schlieren photographs of a bullet penetrating a sheet of glass. One can see the cracks spreading in the glass and can also see the stress waves moving in the glass ahead of the cracks. (*Reproduced by courtesy of the* Laboratoire Central de l'Armement, Paris, *and of* '*J. Soc. Mot. Pict. Telev. Engrs*'.) (Fayolle and Naslin 1954.)

different viewpoint, and on this account the method is impracticable for magnifications greater than unity. However, for remote objects parallax effects are negligible, and Professor Schardin (Schardin 1957 a) considers that this method then produces the best pictures for self-luminous phenomena, though for anything more than a few frames the equipment becomes very expensive (A.R.D.E. 1957).

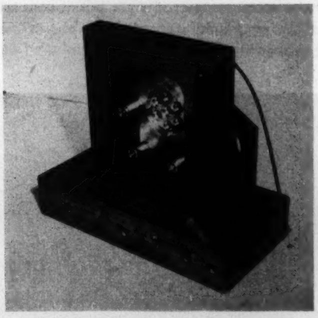


Figure 31. A simple camera that can be used for Cranz–Schardin photography. There is a ring of six camera lenses, and illumination could be provided by six sparks. It has been more convenient to use a cathode-ray tube with a circular time base. A mechanical capping shutter is provided. The central (seventh) lens can conveniently be used for a photograph of the initial layout. (*Reproduced by courtesy of* Messrs. Dunod et Cie, Paris.)

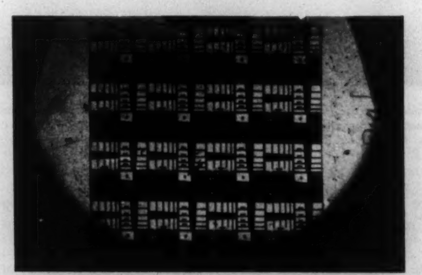


Figure 32. One of the six resolution charts photographed with the 6/7 lens camera shown in the last figure, using cathode-ray tube illumination with an exposure time between 1 and 2 microseconds. (*Reproduced by courtesy of* D. P. C. Thackeray, *and* Dunod et Cie, Paris.)

(*g*) For objects that are not self-luminous and that can be photographed in silhouette, much simpler and more elegant methods are those developed by Cranz and Schardin (Christie 1956, Courtney-Pratt and Thackeray 1956 a, b, Cranz and Schardin 1929, Haensel and Schardin 1956, Hills 1957, Schardin 1954, 1956, 1957 c, Schardin and Struth 1937). Separate cameras or camera lenses are used for each frame and separate light sources are so arranged that though each illuminates the object, light from any one source exposes the emulsion only in the corresponding camera (see figure 29). The light sources may be fired at any chosen intervals and a corresponding series of pictures obtained. Interframe intervals and exposure times can both be kept to fractions of a microsecond. The quality of the pictures can be high as it need only be limited by the performance of the separate camera lenses. Schlieren pictures are possible as well as simple silhouette pictures, and a typical series of schlieren photographs of a bullet passing through a sheet of glass are shown in figure 30.

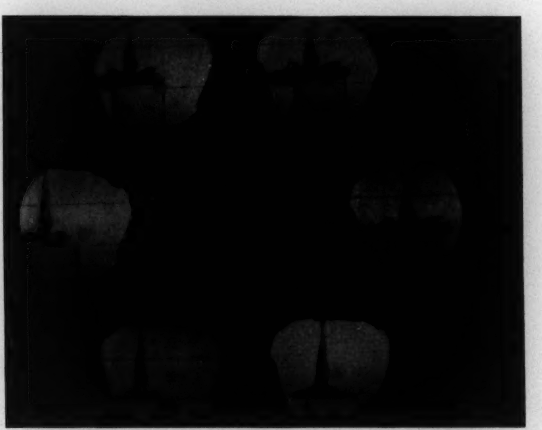


Figure 33. Six successive views, also taken with the 6/7 lens camera and cathode-ray tube illumination, of a fast-moving jet of fluid. Exposure time for each was 2×10^{-6} sec ; interval between frames was 6×10^{-6} sec. The order is clockwise from the bottom right. (*Reproduced by courtesy of* J. Brunton.)

About three years ago we showed that the light source arrangements may be considerably simplified by substituting a single cathode-ray tube for the whole array of spark sources (Courtney-Pratt and Thackeray 1956 a). Our camera is shown in figure 31, a resolution chart in figure 32, and a series of 2-microsecond exposures of a rapidly moving jet of water in figure 33.

(*h*) As described in the preceding paragraph, Cranz and Schardin, in 1929, developed a technique for taking a short series of photographs of objects in silhouette. A severe limitation of their technique is that the photographs of the object are smaller than full size. Not long ago we devised an image splitting optical system which would overcome this disadvantage (Courtney-Pratt and Thackeray 1956 a, b, 1957). A single objective is used and it can conveniently be a microscope objective of almost any focal length (refer to figure 34). Light from a small source at a position S_1 illuminates the object in transmission or silhouette, falls on the objective and is focused to a real image of the source at a position just behind the objective. A small mirror placed at or near this focus can deflect the beam of light in any convenient direction. A real enlarged image of the object is formed at a position I_1. If the source were moved to a position S_5, the real image of the source would be displaced, and a second small mirror could deflect the beam so that the real image I_5 of the object is widely separated from the position I_1. With a ring of spark sources and a pyramidal array of reflectors a number of pictures can be taken in sequence. A cathode-ray tube could be used instead of the ring of spark sources. The system is good for photography at magnifications up to 200 ×. It suffers from the limitation that it is not suitable for the photography of objects by reflected light nor of diffusely transmitting objects, nor of self-luminous objects. The instrument can take a sequence of eight pictures with exposure time for each frame of 10^{-6} sec, and the interval between frames 10^{-6} sec or anything longer. There are no moving parts and the instrument is free from vibration. There are several ways in which the number of pictures can be increased. The resolution with this optical system cannot be as high as if the illumination were arranged to make use of the whole aperture of the objective for each frame. However, the resolution is better than is usually achieved in high-speed systems. Approximately, the resolution is half the best that can be achieved under static conditions with a good microscope.

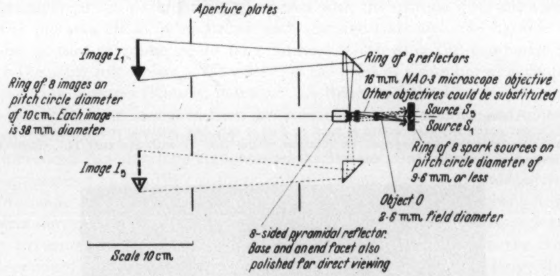

Aperture plates
Image I_1
Ring of 8 reflectors
16 mm NA 0·3 microscope objective
Other objectives could be substituted
Ring of 8 images on pitch circle diameter of 10 cm. Each image is 38 mm diameter
Source S_5
Source S_1
Ring of 8 spark sources on pitch circle diameter of 9·6 mm or less
Image I_5
Object O
2·6 mm field diameter
8-sided pyramidal reflector. Base and an end facet also polished for direct viewing
Scale 10 cm

Figure 34. A diagram of the optical system of the image splitting silhouette microscope. An image of the source S_1 falls on a facet of the pyramidal reflector. When the source S_1 is fired the image I_1 is recorded. (*Reproduced by courtesy of* Butterworths Scientific Publications.)

There is another development of ideas used in a number of the cameras above (Courtney-Pratt 1957). The object is viewed for all pictures of the sequence through one objective. The facets of a pyramidal reflector are used as shown in figure 35 to split the image from the various parts of the objective. Sequential recording is effected by means of a slot or hole in a small rotating disc between the objective and the pyramidal reflector. Figure 36 shows a sequence of pictures taken in this way. As the disc is small it can be rotated at high angular speed and a rapid sequence of pictures obtained. The disadvantages of inclusion of a moving element are minimized as it is small and light. The microscope is capable of taking a sequence of 8 or 16 pictures of objects illuminated in transmission or reflection, bright field, dark field or oblique lighting. It can also be used for self-luminous objects. Useful magnifications range up to 500 × ; the instrument can also be used for the photography of remote objects. Rates of about 10^5/sec can at present be achieved, and with some development higher rates would be possible. The simple instrument can be converted in a few minutes for use either as described here or for silhouette photography with a series of spark sources as described above.

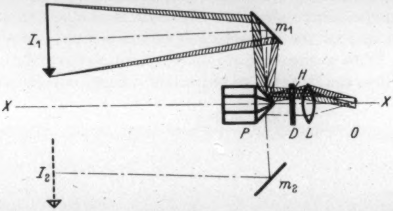


Figure 35. A diagram of the optics of the rotor disc cine-microscope. The light passing through different parts of the objective is split so that it can give separate images such as I_1, I_2, and these are successively exposed by rotation of the disc D with aperture H. (*Reproduced by courtesy of* Butterworths Scientific Publications.)

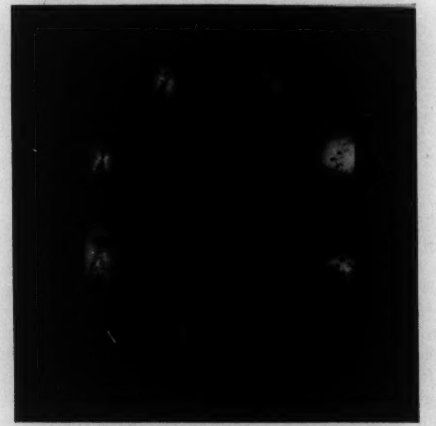


Figure 36. A sequence of 8 pictures taken at 15 000/sec at a magnification from object to plate of 7 ×, using the rotor disc image splitting cine-microscope. The sequence shows the fusing of an electron microscope grid by a surge of current. The illumination was provided by a rectangular pulse of 1400 A through a type FA.5 xenon flash tube for a duration of 6×10^{-4} sec. The order is clockwise from the bottom right. (*Reproduced by courtesy of* Butterworths Scientific Publications.)

§6. SERIES OF PICTURES BY IMAGE DISSECTION

To obtain a series of pictures by image dissection the whole of any one picture or frame is recorded at the same time. Adjacent parts of the picture are not recorded contiguously, but are spaced out so that corresponding parts of subsequent pictures can be recorded between them. The composite record in general will not look like a picture of the object. The individual pictures can be reconstituted from the composite record.

(*a*) A picture could be simply dissected into line elements by photographing the object through a ruled plate placed close in front of the photographic emulsion. If the transparent lines are narrow compared with the opaque intervals a number of similar pictures could be recorded with the line elements side by side. Any particular picture or frame could be examined by viewing the composite record through the same ruled plate. This formed the basis of a number of early patents for changeable pictures (Kanolt, patents ; La Reliephotographie, patents a). For a long time their use seems to have been restricted to novelties and advertising displays. Fordyce Tuttle realized the possibilities of this system, studied its limitations, and has developed apparatus that is fast but simple. In one form of his apparatus (Tuttle 1949 a) there is a dissecting plate with transparent lines 0·001 in. wide at 0·030 in. pitch. The photographic plate is given a sudden transverse movement relative to the dissection plate by a simple spring. A capping shutter prevents double exposure. Assuming no image spread in the emulsion 30 independent pictures could be recorded in the time it takes to move the plate one pitch distance. In his second apparatus (Tuttle 1949 b) he dissects the picture into point elements using a disc pierced with holes of diameter 0·0005 in. at 0·015 in. pitch. The arrays of holes are arranged in arithmetic spirals so that they form a multiple Nipkow disc. This disc is rotated just in front of the emulsion and if there were no image spread it would be possible to record 900 independent pictures. In addition to the image spread due to the emulsion there will be penumbra and diffraction effects increasing with the aperture of the lens and the spacing between the emulsion and the disc. With reasonable figures for these parameters the image width at half density would be at least 0·0013 in., so that the number of independent pictures reduces to 200 or so. Even so, the number of pictures per second can easily be higher than 2×10^5, though the manufacturing tolerances are formidable.

(*b*) O'Brien and Milne (1949) dissected a picture into 30 strip elements using effectively individual lens trains for each element. The elements were assembled end to end across film mounted on a rotating drum. Although the resolution was very poor the rate at which pictures could be taken approached 10^7 per second. In their most recent designs even this speed has been surpassed, and the resolution has been increased to 48 lines across the frame.

(*c*) Sultanoff (1950 a, b, 1954) used a ruled plate to dissect the picture and a rotating mirror camera to sweep the image over the photographic emulsion, as shown in figure 37. He has taken pictures at rates of the order of 10^8/sec. The dissection plate was ruled with 250 lines 0·0005 in. wide spaced 0·015 in. apart. The main camera lens formed an image of the object on this dissection plate. A second lens collected light that came through the dissection plate and formed an image at unity magnification on a 5 in. × 4 in. photographic plate after reflection from the rotatable mirror. The sweep speed of the image could be 3000 m/sec. Were there no image spread 30 pictures could be recorded at 10^8 pictures per second, the numerical aperture could be f/9, and the resolution 250 lines across the field. In practice there is image spread due to the emulsion and due to imperfections in the lens and mirror. Moreover, any distortion introduced by this optical system will mean that, in unscrambling, parts of a picture recorded at some instant will appear in some subsequent or earlier picture. The image spread introduced by the emulsion or by the optical system will reduce the number of independent pictures that can be recorded, the rate at which they may be taken, and the effective numerical aperture. Synchronization may be a problem as the recording must take place at precisely the right instant with respect to the rotation of the mirror.

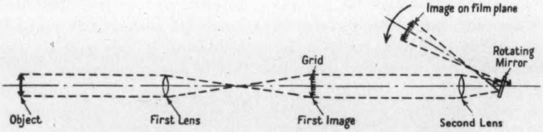


Figure 37. Schematic diagram of the camera designed by Morton Sultanoff. An image of the event falls on a ruled grid ; light passing through this dissection plate is imaged after reflection from a rotating mirror on the photographic emulsion. The successive dissected pictures are recorded with their line elements side by side. They can be subsequently unscrambled by viewing in the camera, or by viewing through an identical dissection plate held in contact with the composite negative.

It is interesting to note that with a design like this the product of the number of lines resolved across the frame and the framing rate could ideally be some 250 times higher than the Schardin limit for a simple rotating mirror camera. Sultanoff's design is nowhere near its ultimate limit but does surpass the Schardin limit by a considerable margin, and has already proved of great value for really high-speed studies of brilliantly illuminated events.

Other workers (Armament Research Establishment 1952 b, Uyemura and Morishige 1957) have used equipment similar in principle to Sultanoff's, but much simpler in that the dissection plates had very few slits (usually seven). This greatly eases the stringency of the optical and mechanical tolerances.

(*d*) Dissection by a block of light guides is possible. Young and Legg of Reyrolle's have used a block of about 30 × 30 light guides to dissect an image formed by a camera lens (Young and Legg 1955). The remote ends of all these guides have been rearranged in a line across the film. Each guide traces on the film a line of density corresponding to the intensity of light falling in the particular part of the image. This scheme allows rapid recording—up to about 10^5 frames per second—and in addition allows the recording of very long sequences of pictures, though they are of poor spatial resolution because of the limited number of guides that can be aligned across the film. Some further work on systems of this general sort is under way in this country and at the Institute of Optics at Rochester, N.Y., in America.

(*e*) In the Research Laboratory for the Physics and Chemistry of Surfaces I have developed a number of aperture scanning image dissection cameras (Courtney-Pratt 1953 a, b, Courtney-Pratt and Thackeray 1956 b). The record is made on a composite plate which later can yield a sequence of dissected two-dimensional pictures. A lenticular plate with an array of 'spherical' lenslets is used to dissect the image into elements which are moved over the photographic emulsion by altering the direction in which light falls on the plate. To be more precise, two plates are used, each embossed on one surface with a parallel array of cylindrical lenslets at a pitch spacing of 25/cm. A pair of these plates mounted face to face with axes crossed at right angles act as if the combination were a plate embossed with an array of 625 'spherical' or point-focus lenslets per square centimetre (La Reliephotographie, patents b). Such plates, similar to those required for stereoscopic photography by the Bonnet process (La Reliephotographie, patents b, c, Butement 1948, Judge 1950), have been supplied by Monsieur R. Marilhet, Director of *Photographie en Relief*, 152, Av. des Champs Elysées, Paris. The method has proved valuable in a whole range of photographic problems, and a number of cameras have been built for particular purposes. The optical system is shown in figure 38. In the simplest camera, sequential recording is effected by the rotation of a Nipkow disc between the components of the camera lens; 200 pictures can be recorded at rates up to 50 000 per second. In each picture there are 40 000 picture elements, a resolution of 200 lines across the field in each dimension. The effective numerical aperture is f/6·3. Figure 39 is a series of 25 frames from the 200 recorded when the camera was used to study the explosion of mercury fulminate. This was taken during part of a general research programme in our laboratory on the mechanism of explosion in liquids and solids (Bowden and Yoffe 1952, Evans and Yoffe 1956).

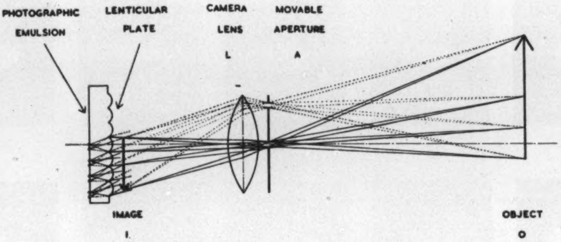


Figure 38. Diagram of the optical layout for lenticular-plate image-dissection aperture-scanning multiple-frame photography. Solid lines indicate the light paths when the aperture is in its central position; broken lines indicate the light paths when the aperture is in a peripheral position. (*Reproduced by courtesy of the 'Journal of Photographic Science'.*)

A similar camera, shown in figure 40, of overall dimensions 10 cm × 9 cm × 8 cm, can take 100 pictures at 80 000 per second with a resolution of 90 lines across the field in both dimensions. Lenticular plate Nipkow disc cameras working on these principles are now being manufactured by J. Langham Thompson Ltd.,

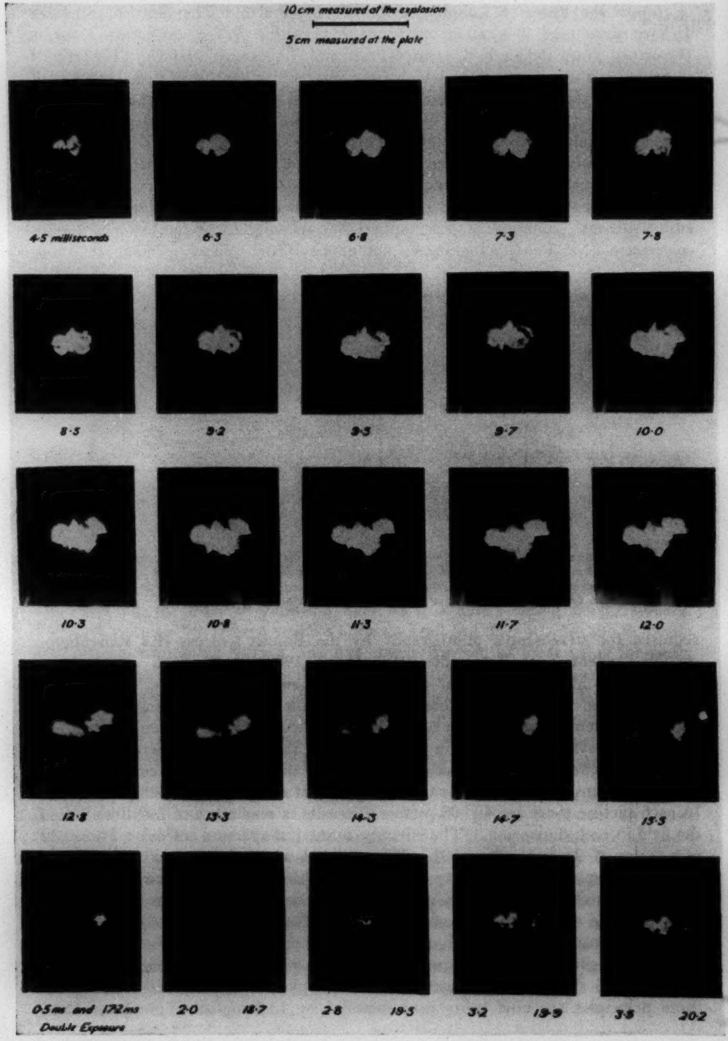

10 cm measured at the explosion
5 cm measured at the plate
4·5 milliseconds
6·3
6·8
7·3
7·8
8·5
9·2
9·5
9·7
10·0
10·3
10·8
11·3
11·7
12·0
12·8
13·3
14·3
14·7
15·3
0·5 ms and 17·2 ms
Double Exposure
2·0
18·7
2·8
19·5
3·2
19·9
3·5
20·2

Figure 39. Twenty-five prints selected from the 200 recorded showing stages in the growth of the explosion of some mercury fulminate. Times measured in milliseconds from the commencement of the explosion are given beneath the pictures. (*Reproduced by courtesy of the 'Journal of Photographic Science'*.)

Springland Laboratories, Bushey Heath, Hertfordshire (Hayes 1955, 1956, J. Langham Thompson Ltd. 1957). These cameras, one of which is shown in figure 41, can take sequences of 300 pictures of resolution 250 × 300 lines at any desired rate between 100 pictures per second and 250 000 pictures per second. Figure 42 is a typical series. All of these cameras can be used for the photography of self-luminous objects or of objects separately illuminated.

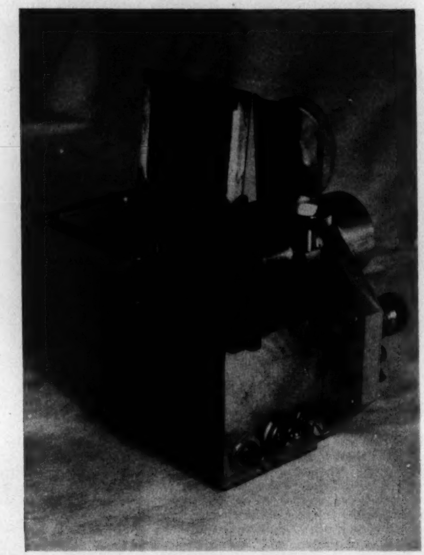


Figure 40. Miniature high-speed camera, of overall dimensions 10 cm × 9 cm × 8 cm. Sequential recording is effected by means of an enclosed Nipkow disc. The camera can take 100 pictures at rates up to 80 000/sec. On the rear of the plate holder are hinged frames that hold a ground-glass screen for viewing, and a pressure plate for use when making unscrambled contact prints. Speed controls, flash-lamp switch, geared speed indicator, and focusing controls can be seen. A capping shutter with manual or automatic settings is provided. (Designed by J. S. Courtney-Pratt. Photographed by G. J. Dean. Manufactured by R. C. Moss, C. A. Naunton and A. Priddle.)

For transparence or silhouette photography, no Nipkow disc is necessary, and in some of the cameras for such work mechanically moving parts have been entirely eliminated. Separation of the individual frames is then achieved by movement of the light source. In one apparatus the source of light is the luminous spot of a cathode-ray tube. Series of 200 pictures are recorded at 200 000 per second. The exposure time and the rate of taking pictures can be independently controlled. Synchronization with the phenomenon is simple.

In another apparatus a short trail of lead azide was used as a moving source of light. Twenty pictures were recorded at intervals of a quarter of a microsecond.

It is possible to dispense with the field and camera lenses. The reconstituted pictures are each similar to ordinary spark shadowgraphs. With a cathode-ray tube as the light source, as in figure 43, series of 200 pictures of high resolution can be recorded at 10^5 to 10^6 pictures per second. With lead azide as the light source series of 20 pictures can be recorded somewhat faster. It is perhaps remarkable that photography is possible at such a rate in this simple apparatus in which there are no moving parts, no electronic devices, and no camera lenses.

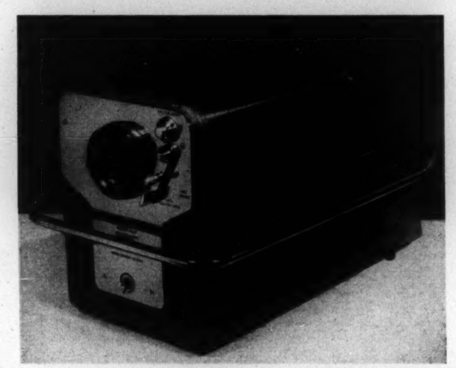
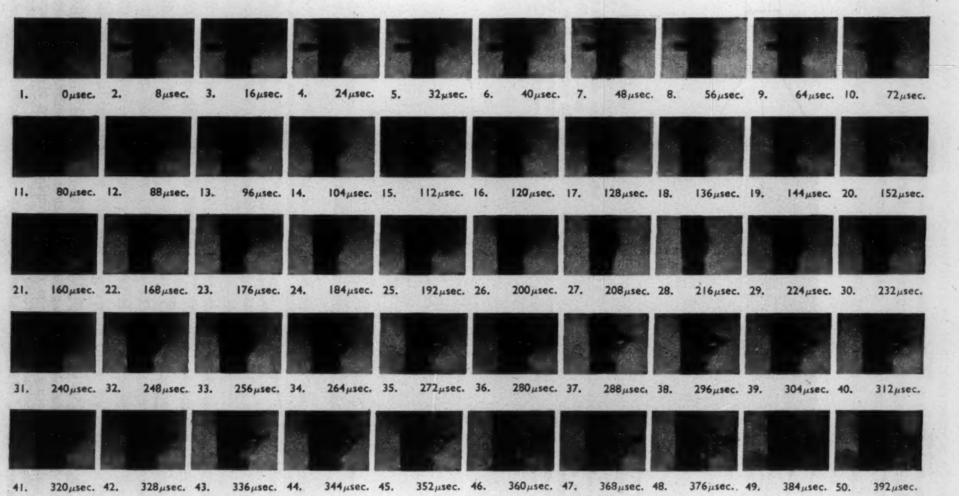


Figure 41. The C.P. Series 600 camera manufactured by J. Langham Thompson Limited. This camera uses a lenticular plate to dissect the picture into 80 000 elements (i.e. about 250 × 300 lines across the frame). Sequential recording is effected by means of a Nipkow disc rotating in the aperture plane of the main lens. It can take 300 pictures at any rate from 100 pictures per second to 250 000 per second. The effective numerical aperture is about f/6, though in certain applications the image may be two or three times brighter than would be achieved with a conventional system of this numerical aperture. (*Reproduced by courtesy of* J. Langham Thompson Limited, R. A. Hayes, *and ' Industrial Image '*.)

All of these cameras can be simply modified for stereoscopic recording though then the number of independent pictures of a series is halved. The cameras can all be used to unscramble the composite record for direct slow-motion viewing and so that individual pictures may be examined in detail.

There are a number of further developments. For example, J. L. Viard has introduced the idea of ' virtual ' aperture scanning by the use in one case of a rotating mirror, and in another by the use of a row of small Kerr cells. These methods will allow a considerable increase in speed (Viard 1957).

(*f*) I have applied similar principles of lenticular plate dissection to cine-micrography, so that series of micrographs can be recorded at high speed (Courtney-Pratt 1956 a, Courtney-Pratt and Thackeray 1956 b). Instead of allowing the enlarged image to fall directly on a photographic emulsion, it is made to fall, as in figure 44, on a plate embossed with a large number of small



1. 0μsec. 2. 8μsec. 3. 16μsec. 4. 24μsec. 5. 32μsec. 6. 40μsec. 7. 48μsec. 8. 56μsec. 9. 64μsec. 10. 72μsec.
11. 80μsec. 12. 88μsec. 13. 96μsec. 14. 104μsec. 15. 112μsec. 16. 120μsec. 17. 128μsec. 18. 136μsec. 19. 144μsec. 20. 152μsec.
21. 160μsec. 22. 168μsec. 23. 176μsec. 24. 184μsec. 25. 192μsec. 26. 200μsec. 27. 208μsec. 28. 216μsec. 29. 224μsec. 30. 232μsec.
31. 240μsec. 32. 248μsec. 33. 256μsec. 34. 264μsec. 35. 272μsec. 36. 280μsec. 37. 288μsec. 38. 296μsec. 39. 304μsec. 40. 312μsec.
41. 320μsec. 42. 328μsec. 43. 336μsec. 44. 344μsec. 45. 352μsec. 46. 360μsec. 47. 368μsec. 48. 376μsec. 49. 384μsec. 50. 392μsec.

Figure 42. A sequence of pictures taken with the C.P. Series 600 camera manufactured by J. Langham Thompson Limited. It shows a ·45 revolver bullet passing through a block of wood. The speed of the bullet is 650 ft/sec, and the effective framing rate is 125 000 pictures per second. (*Reproduced by courtesy of* J. Langham Thompson Limited, R. A. Hayes, *and 'Industrial Image'*.)

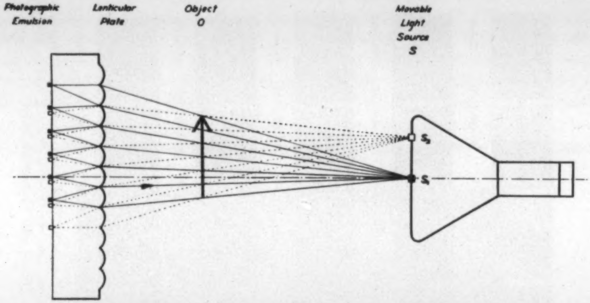

Photographic Emulsion
Lenticular Plate
Object O
Movable Light Source S
S_2
S_1

Figure 43. A diagram to indicate the way that image dissection may be used to take a sequence of shadowgraphs. A source S_1 casts a shadow of the object O on the lenticular plate. A second source S_2 could similarly cast a shadow on the lenticular plate and the image elements would be displaced and distinct from the first set. Spark sources can be used. Alternatively, the luminous spot of a cathode-ray tube provides a convenient movable source. It is possible to use a detonating trail of lead azide to record a limited series of pictures at rates of several million per second. (*Reproduced by courtesy of the 'Journal of Photographic Science'.*)

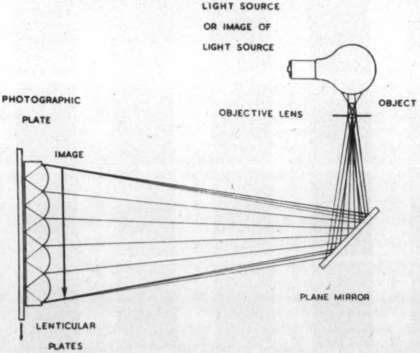

LIGHT SOURCE OR IMAGE OF LIGHT SOURCE
PHOTOGRAPHIC PLATE
OBJECTIVE LENS
OBJECT
IMAGE
PLANE MIRROR
LENTICULAR PLATES

Figure 44. Simplified diagram of the optical layout for the lenticular plate image dissection cinemicrograph. The lens forms a real image of the object on the lenticular plate. Each lenslet of the lenticular plate forms an image on the photographic emulsion of the aperture of the microscope objective. Sequential recording is effected by traversing the photographic plate. (*Reproduced by courtesy of* Messrs. Dunod et Cie, Paris.)

lenses. Each of these lenslets focuses the light that falls on it to a small dot (which is actually an image of the aperture of the microscope objective). These small dot elements are all recorded simultaneously on a plate, but each element is separated from its neighbours by a distance that is large compared with the size of the element. If the photographic plate is moved sidewards by the width of one dot a new set of picture elements can be recorded beside the last set. That is, a second picture has been recorded interlaced with the first, and the displacement required needed only to be about 1/650 of a centimetre instead of the width of a frame. If the plate is traversed at 20 centimetres per second one could record pictures in this way at $20 \times 650 = 13\ 000$ pictures per second.

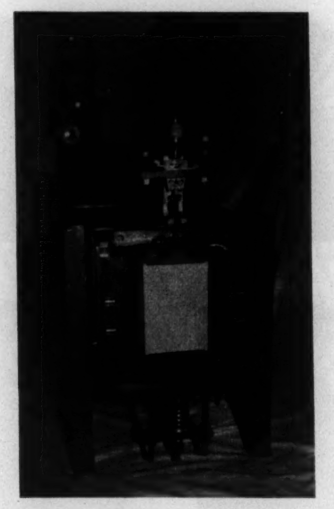
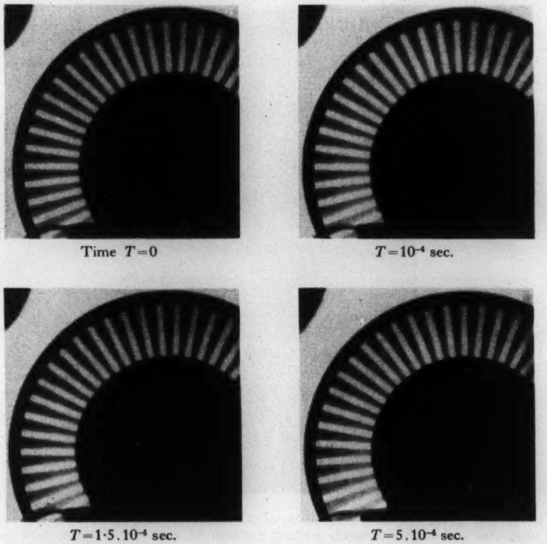


Figure 45. A view of the operating mechanism of the lenticular plate cine-micrograph. The plate is supported and guided vertically without backlash on a system of elastic steel links. The toggle-catch mechanism at the top is arranged to give rapid and reproducible release. The additional accelerator spring and buffers which can be seen at the bottom of the figure can be connected to give higher velocities. (*Reproduced by courtesy of* Messrs. Dunod et Cie, Paris.)

If the direction of traverse were parallel to the axes of the lenticular plates, it would be possible to move the photographic plate by a distance equal to one pitch spacing of the lenslets before overlapping of images and double exposure occurred. If, however, the direction of traverse makes an angle of 1 in 13 with the axis of the lenticular plate, it is possible to traverse the photographic plate 13 times as far before double exposure would occur. Using Ilford N40 process plates it is possible to resolve thirteen clearly separated dots per pitch distance behind each lenslet. This means that without overlapping one could distinguish

13^2 complete alternations of intensity without double exposure. That is, the equipment can take a sequence of $2 . 13^2 = 338$ separate frames in a sequence.

The working parts of the cine-micrograph are shown in figure 45. If the plate holder is drawn 1 cm away from centre and then released it is accelerated by the stored energy of the elastic suspension system. The velocity in mid-swing is 32 cm/sec. Additional accelerating springs are provided so that this may be increased by a factor of 2½. It is thus easy to see that the equipment can take pictures at rates of 50 000/sec. Parts of a Vickers Projection Microscope have been used in the construction of this apparatus, and this makes it easy to use the equipment for micrography with any conventional type of illumination—transmitted light, reflected light, bright field, dark field, phase contrast, schlieren, etc.



Time $T=0$

$T=10^{-4}$ sec.

$T=1{\cdot}5 . 10^{-4}$ sec.

$T=5 . 10^{-4}$ sec.

Figure 46. Four prints from a series of 300 recorded on one composite plate with the cine-micrograph shown in the last figure. In playback the slotted rotor and the timing wheel can be seen to rotate smoothly. The time resolution is considerably better than 10^{-4} sec. (*Reproduced by courtesy of* Messrs. Dunod et Cie, Paris.)

The distance from the objective (or projection eyepiece) to the lenticular plate is intentionally long. This allows the use of a large plate within the narrow field angle of ordinary microscope optics. A large lenticular plate gives a higher picture resolution. The plates in this microscope are 5 inches square, and the resolution is 320 lines across the field in both dimensions. Further, as this distance is long, it is not necessary to stop down the objective in any fashion and one can use microscope objectives of as high numerical aperture as are available. This leads to a high specific brilliance in the image at the photographic emulsion, as all the light collected by the microscope objective and falling on any one lenslet of the lenticular plate is concentrated to a patch about 600 times smaller than the area of the lenslet. This advantage is of major importance, as normally it is the available light level which limits the speed and magnification in cine-micrography.

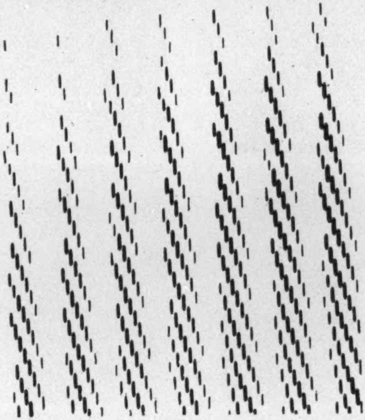


Figure 47. An enlargement at 25 × of part of the composite plate on which was recorded the movement of the slotted rotor illustrated in the last figure. The vertical trace behind each lenslet is alternately dark and light depending on whether or not there was falling on it the image of a spoke or a slot of the rotor. (*Reproduced by courtesy of* Messrs. Dunod et Cie, Paris.)

With this apparatus it is possible to use one ordinary magnesium–oxygen photoflash bulb for taking a whole series of pictures at 50 000 per second at magnifications up to 500 ×, using Ilford N40 process plates. Higher magnifications are possible at slower speeds. The apparatus can also be used at smaller magnifications and, indeed, can be used for the photography of remote objects. Four prints showing the rotation of a slotted wheel are shown in figure 46, and figure 47 is an enlargement of a small section of the composite plate from which the photographs of figure 46 were unscrambled. Figure 48 has six views selected from the complete record of the ignition of a small crystal of mercury fulminate.

Electronic flash sources have been built by Thackeray for use with this microscope (Thackeray 1957). Their specific brilliance is at least 10 times greater than the minimum required to make a recording at 500 ×. They have a light output in the form of a single rectangular pulse of duration long enough for the recording of a large number of frames. The pulse can be precisely synchronized, with a delay not more than a few microseconds.

As the time from the release of the traverse mechanism to maximum traverse speed is relatively long (about 1/20 sec) it is often more suitable to initiate the

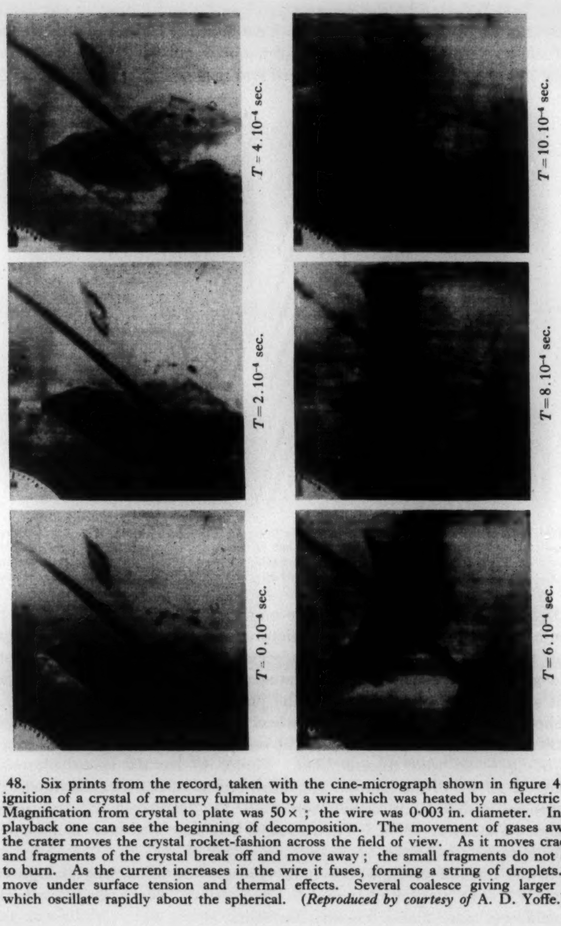


Figure 48. Six prints from the record, taken with the cine-micrograph shown in figure 45, of the ignition of a crystal of mercury fulminate by a wire which was heated by an electric current. Magnification from crystal to plate was 50 × ; the wire was 0·003 in. diameter. In the full playback one can see the beginning of decomposition. The movement of gases away from the crater moves the crystal rocket-fashion across the field of view. As it moves cracks form and fragments of the crystal break off and move away ; the small fragments do not continue to burn. As the current increases in the wire it fuses, forming a string of droplets. These move under surface tension and thermal effects. Several coalesce giving larger droplets which oscillate rapidly about the spherical. (*Reproduced by courtesy of* A. D. Yoffe.)

event to be photographed from the camera than initiate both from some external synchronizing signal. In this respect this type of equipment is less convenient than some other cameras.

In the summer of 1955, I designed a more elaborate apparatus of the same type as this cine-micrograph, and it has been built for use at the General Electric Research Laboratories, The Knolls, Schenectady, N.Y., U.S.A. Speed and resolution have both been about doubled (Courtney-Pratt and Huggins 1956, 1957). This equipment is shown in figures 49 and 50.

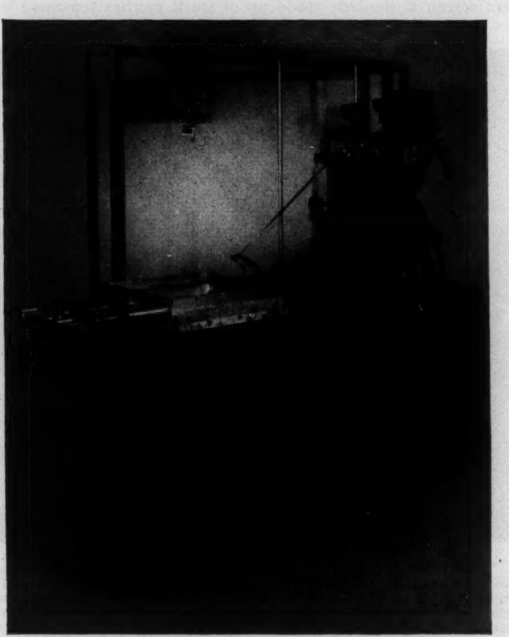


Figure 49. A general view of the latest lenticular plate cine-micrograph. A complete Vickers projection microscope has been incorporated. The lenticular plates in this equipment are 6 in. square, and the picture resolution is 375 lines across the frame each way. It can record pictures at any of a number of speeds up to at least 100 000 pictures per second at magnifications up to 2000 ×. A 16-mm cine-camera is used to copy the unscrambled record so that one may have if desired a duplicate record that can be projected through standard equipment. (*Reproduced by courtesy of the* General Electric Research Laboratories, Schenectady, N.Y., U.S.A.)

A lenticular plate camera with a traversing photographic plate has also been designed and used by G. R. R. Bray for photographing ballistic phenomena (Bray 1956).

(*g*) Recently, G. H. Lunn, working at the Atomic Weapons Research Establishment, Aldermaston, has combined the principles of the image converter and of image dissection (Lunn 1957, Lunn and Chippendale 1957). A dissected cathode was built into an image converter tube, and the cathode then consisted of 0·001 in. squares of photosensitive material at 0·010 in. spacing. The image on the fluorescent screen is similarly made up of small patches separated by relatively large distances. A magnetic time-base can sweep these images across the screen and the composite record on the screen can be recorded by a single wide-aperture lens on a fixed emulsion. As with the lenticular plate cine-micrograph described above, the sweep direction should make a small angle with the array of image elements. An angle of about 1 in 5 is suitable. This allows one to take a sequence of about 50 pictures at very high rates. The maximum rate would appear to be governed mostly by the throughput of light, and it is a major disadvantage of the scheme (except for extremely brilliant subjects) that only 1% of the light forming the primary image in the plane of the cathode falls on photosensitive material.

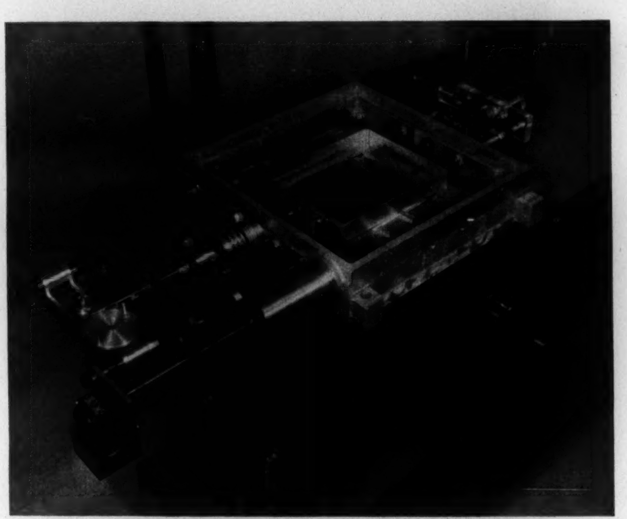


Figure 50. An enlarged view of the operating mechanism of the cine-micrograph shown in the preceding figure. As with the earlier equipment the plate holder is supported on a system of elastic steel links. Inertial masses may be coupled for low-speed recording. Automatic playback and measurement facilities are provided. The lenticular plates are mounted in a rotatable frame that may be fixed by dowels to give various traverse angles so that more efficient use may be made of fine grain slow, or coarser grain fast, emulsions. (*Reproduced by courtesy of the* General Electric Research Laboratories, Schenectady, N.Y., U.S.A.)

Lunn has taken pictures of the growth of a flash discharge, at rates of 5×10^7 per second, and there seems no reason why the equipment should not be developed within a short time to take pictures at 10^9/sec, or, with considerably greater development troubles, at rates a good deal higher still. At these higher speeds the throughput of light will almost certainly be the governing factor.

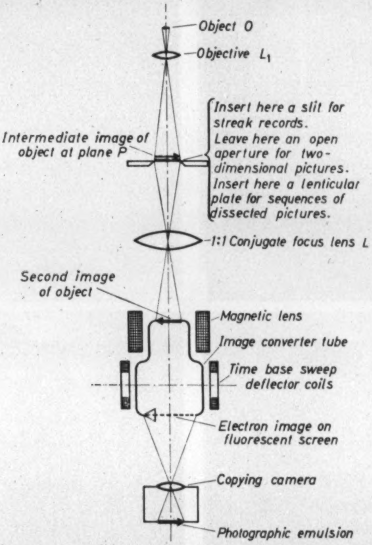

Object O
Objective L_1
Insert here a slit for streak records.
Leave here an open aperture for two-dimensional pictures.
Insert here a lenticular plate for sequences of dissected pictures.
Intermediate image of object at plane P
1:1 Conjugate focus lens L
Second image of object
Magnetic lens
Image converter tube
Time base sweep deflector coils
Electron image on fluorescent screen
Copying camera
Photographic emulsion

Figure 51. Schematic diagram of image converter tube apparatus with an intermediate image plane at P. If a slit is placed at P this apparatus can be used for streak records. If an open aperture is left at P the apparatus can take ordinary two-dimensional pictures of very short exposure. If a lenticular plate is placed at P the image at the photo-cathode is dissected into small dot elements. If a time sweep is then applied to the tube the composite dissected picture sequence appears on the fluorescent screen and can be recorded by a plate camera. The individual pictures of the sequence can subsequently be unscrambled. (*Reproduced by courtesy of* Butterworths Scientific Publications.)

It is possible to use a lenticular plate along with the standard Mullard image converter tube so that in many applications a greater proportion of the incident light is effective (Courtney-Pratt and Thackeray 1956 b). It would be possible to put a lenticular plate just outside the image converter cathode, but the focal length of the lenticular plates that we have is too short, and it is not at the moment possible to bring them close enough to the photosensitive surface. However, we have found that one can use two wide-aperture aircraft-camera lenses, mounted face to face for forming an image at unity magnification. The combination of two 12-in. f/2·5 Aero Ektars will resolve 2000 equally spaced line pairs across a

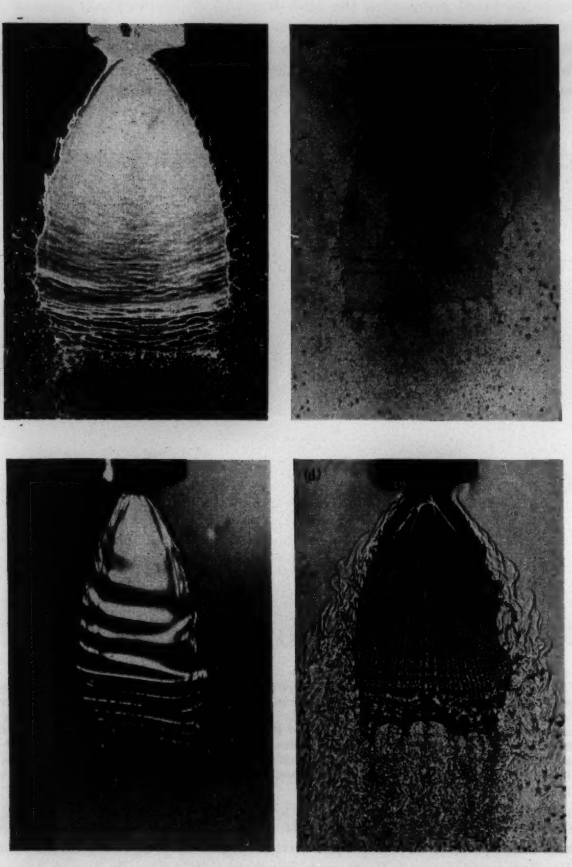


Figure 52. Four photographs showing the disintegration of laminar jets or sheets of liquid. The exposure in each case is one microsecond and the pictures illustrate the wide differences in observable detail when different methods of lighting are used. (*a*) Diffuse reflection, (*b*) diffuse transmission, (*c*) specular reflection, (*d*) parallel transmission. (*c*) and (*d*) were taken simultaneously. (*Reproduced by courtesy of* R. P. Fraser and N. Dombrowski, and Butterworths Scientific Publications.)

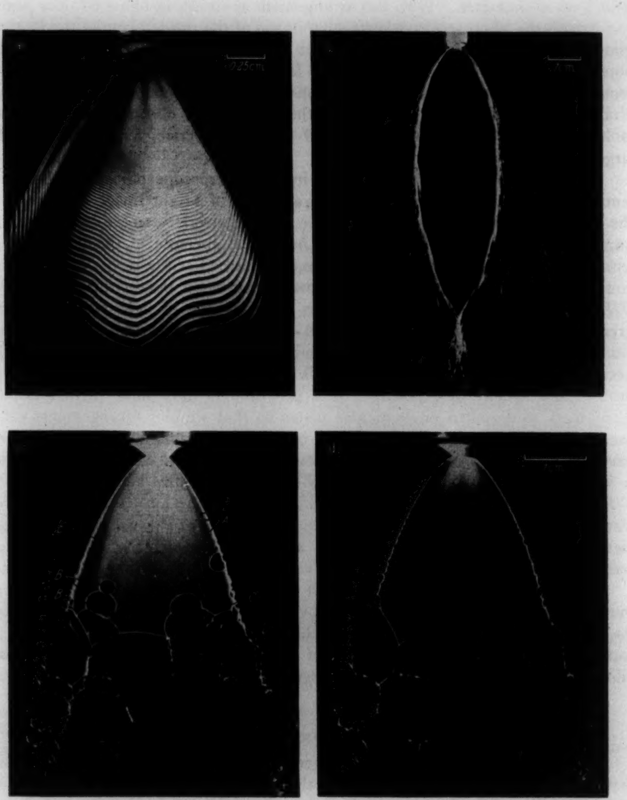


Figure 53. Four further illustrations of the disintegration of laminar liquid sheets. (*a*) An interferogram with an exposure time of 1/100 sec ; (*b*) an exposure of 1/1000 sec ; small particles of metal within the liquid act as point reflectors and show flow direction ; (*c*) and (*d*), microsecond exposures. For (*c*) there were two exposures a short time apart, and though the spectral band of each flash was different, the emulsion would record both. The photographic emulsion for the record (*d*) was placed face to face with the emulsion for the record (*c*) but the emulsion for (*d*) would only respond to the spectral band from one of the two flashes. (***Reproduced by courtesy of*** R. P. Fraser and N. Dombrowski, and Butterworths Scientific Publications.)

field 1 in. in diameter. With this arrangement as shown in figure 51, it is possible to form on the photosensitive surface an image of the image plane of the lenticular plates. If now an image of the event is formed by means of a microscope objective (or even a macrolens) on the lenticular plate, the picture will be dissected into elements which will be small patches provided that the distance V between the lenticular plates and the microscope objective, or other lens, is long compared with the diameter D of that objective, i.e. provided V/D is large.

All the light that can be collected by the microscope objective falls on the lenticular plates and is concentrated to an array of bright image elements. All the light from these is collected by the pair of aircraft-camera lenses, as their numerical aperture is greater than the numerical aperture of the lenslets of the lenticular plate. A bright, dissected image is formed at the cathode of the image converter tube. Though there are some transmission losses, this system when V/D is large (greater, say, than 20) has a throughput of light at least 50 times greater than with the same lens system and a single dissected cathode used by Lunn. This is most important in high-speed micrography, as there V/D is necessarily large and it is almost always the light levels which limit the magnification and brevity of exposure. It is proposed to have some further development work undertaken on lenticular plates so that simpler arrangements can be used. For values of V/D that could be between about 4 and 20 the use of lenticular plates and an arbitrary lens stop can give some improvement. If one is working at low magnification, and wide aperture lenses can be used, so that V/D can be smaller than the numerical aperture of the lenslets of the lenticular plate, there is no advantage on grounds of intensity in the use of the lenticular plate, and indeed the image elements are brighter using the simple dissected cathode.

It is worth noting that the product of lines resolved across the frame and framing speed is already, with image-converter–image-dissection equipment, above the Schardin limit; the system should allow, before fundamental limits are approached, a figure for this product of 10^{13} or 10^{14} sec^{-1}. The inherent possibilities of schemes of this sort are very great indeed.

§ 7. CONCLUSION

In spite of the complexity of some sorts of camera, only too often one finds that the actual use of the camera and the recording is only a small part of the problem. The difficulties of synchronizing and of lighting are often much more severe than the development and use of the camera equipment. Figures 52 and 53 illustrate very clearly the great differences in the fine structure that may be observed with different methods of lighting, and by attention to the detail of the optical layout (Dombrowski 1956, Fraser and Dombrowski 1957).

It is usually well worth going to a more complicated form of photography if thereby the synchronization and/or the lighting requirements are eased. In almost all camera designs it is therefore of major importance to consider the throughput of light and the synchronization problems, as well as the maximum speed and resolution, important as these are.

A wide range of cameras has been developed and it is possible to obtain reasonable photographs of most macroscopic phenomena. The problems are more acute when one wishes to study fine detail, and in high-speed cine-micrography one must make use of the most advanced and elaborate techniques.

The new methods that have been developed can contribute a great deal in research and in industry in fields as widely separated as ballistics, biology, fluid dynamics, electric discharge studies, vibration, fracture, atomization, combustion, metallography, and micrography.

REFERENCES

Note: The abbreviations HSP2, HSP3 are used throughout the list of references for *Actes du 2ᵉ Congrès International de Photographie et Cinématographie Ultra-rapides* (Paris: Dunod, 1954/56), and *The Proceedings of the Third International Congress on High-speed Photography* (London: Butterworths Scientific Publications, 1956/57) respectively.

ADAMS, C. A., 1950, *Proc. Roy. Soc.* A, **204**, 19.

ADAMS, G. K., 1951, *J. Sci. Instrum.*, **28**, 379.

ALDINGTON, J. N., and MEADOWCROFT, A. J., 1948, *J. Instn Elect. Engrs*, Pt. II, **95**, 671.

AMIOT, J. L., Rue des Douaniers, Paris, Commercial Literature.

ARMAMENT RESEARCH ESTABLISHMENT, 1951, *Handbook of the 35th Physical Society Exhibition* (London: Physical Society), p. 177; 1952 a, *Handbook of the 36th Physical Society Exhibition* (London: Physical Society), p. 175; 1952 b, *Ibid.*, p. 176.

ARMAMENT RESEARCH AND DEVELOPMENT ESTABLISHMENT, 1957, *Handbook of the 41st Physical Society Exhibition* (London: Physical Society), p. 7.

AUMONT, R., and VODAR, B., 1956, HSP2, p. 39.

BAIRD, K. M., 1946, *Canad. J. Res.* A, **24**, 41.

BAIRD, K. M., and DURIE, D. S. L., 1949, *J. Soc. Mot. Pict. Telev. Engrs*, **53**, 489.

BEAMS, J. W., 1937, *J. Appl. Phys.*, **8**, 795.

BEAMS, J. W., KUHLTHAU, A. R., LAPSLEY, A. C., MCQUEEN, J. H., SNODDY, L. B., and WHITEHEAD, W. D., 1947, *J. Opt. Soc. Amer.*, **37**, 868.

BECKMAN and WHITLEY INC., San Carlos, California, U.S.A., Commercial Literature.

BOON, J. L., 1944, *J. Soc. Mot. Pict. Telev. Engrs*, **43**, 321.

BOURNE, H. K., 1948, *Discharge Lamps for Photography and Projection* (London: Chapman and Hall).

BOWDEN, F. P., and YOFFE, A. D., 1952, *Initiation and Growth of Explosions in Liquids and Solids* (Cambridge: University Press).

BOYS, C. V., 1893, *Nature, Lond.*, **47**, 415, 440.

BRAY, G. R. R., 1956, HSP2, p. 175.

BRIXNER, B., 1954, *High Speed Photography*, **5**, 55 (New York: Soc. Mot. Pict. Telev. Engrs); 1956, HSP2, p. 108; 1957, HSP3, p. 289.

BROOKES, A. M. P., and MONRO, P. A. G., 1957, HSP3, p. 142.

BUTEMENT, C., 1948, *Sci. News, Harmondsworth*, **9**, 62.

CADY, W. M., and ZAREM, A. M., 1948, *Nature, Lond.*, **162**, 528.

CAIRNS, R. W., 1944, *Industr. Engng Chem.*, **36**, 79.

CHESTERMAN, W. D., 1951, *The Photographic Study of Rapid Events* (Oxford: Clarendon Press).

CHESTERMAN, W. D., and GLEGG, D. R., 1956, HSP2, p. 8.

CHESTERMAN, W. D., and MYERS, D. T., 1951, *J. Sci. Instrum.*, **28**, 301.

CHESTERMAN, W. D., and PECK, G. T., 1952, *Photogr. J.*, **92**B, 133.

CHIPPENDALE, R. A., 1952, *Photogr. J.*, **92**B, 149; 1957, HSP3, p. 116.

CHRISTIE, D. G., 1956, HSP2, p. 327.

COLLINS, R. B., 1957, HSP3, particularly NASLIN, P., p. 1; COURTNEY-PRATT, J. S., p. 403.

CONN, W. M., 1951, *J. Opt. Soc. Amer.*, **41**, 445.

Courtney-Pratt, J. S., 1949, *Research, Lond.*, **2**, 287 ; 1950, *Proc. Roy. Soc.* A, **204**, 27 ; 1951, Brit. Pat. No. 662065 ; 1952, *Photogr. J.*, **92**B, 137 ; 1953 a, *J. Photogr. Sci.*, **1**, 21 ; 1953 b, *Fourth Symposium on Combustion* (Baltimore : The Williams and Wilkins Co.), p. 508 ; 1956 a, HSP2, p. 152 ; 1956 b, *Nature, Lond.*, **178**, 1440 ; 1957, HSP3, p. 87.

Courtney-Pratt, J. S., and Huggins, C. M., 1956, *Gen. Elect. Res. Lab. Rep.*, No. 56RL1621 ; 1957, *Rev. Sci. Instrum.*, **28**, 256.

Courtney-Pratt, J. S., and Thackeray, D. P. C., 1956 a, HSP2, p. 163 ; 1956 b, *Apparatus for High Speed Photography* (London : Butterworths Scientific Publications), see also 1957, *J. Photogr. Sci.*, **5**, 32 ; 1957, HSP3, p. 81.

Cranz, C., and Schardin, H., 1929, *Z. Phys.*, **56**, 147.

Croney, D., 1948, *Nature, Lond.*, **162**, 490.

Devaux, P., 1956, HSP2, p. 43.

Dombrowski, N., 1956, *Chem. Engr*, **34**, 35.

Eastman Kodak Co., Rochester, N.Y., U.S.A., Commercial Literature.

Edgerton, H. E., 1946, *J. Opt. Soc. Amer.*, **36**, 390 ; 1952, *Rev. Sci. Instrum.*, **23**, 532 ; 1957, HSP3, p. 51.

Edgerton, H. E., and Killian, J. R., 1939/54, *Flash !* (Boston : Hale, Cushman and Flint, 1939 ; Branford, 1954).

Edgerton, H. E., and Wyckoff, C. W., 1951, *J. Soc. Mot. Pict. Telev. Engrs*, **56**, 398.

Evans, B. L., and Yoffe, A. D., 1956, *Proc. Roy. Soc.* A, **238**, 325.

Eyles, E. D., 1941, *J. Sci. Instrum.*, **18**, 175.

Fayolle, P., and Naslin, P., 1948, *Mémorial de l'Artillerie française*, **22**, 657 ; 1949, *Ibid.*, **23**, 1. Reprinted together as ' Photographie Instantanée et Cinématographie Ultra-rapide ' in the *Editions de la Revue d'Optique* ; 1954, *High Speed Photography*, **5**, 101 (New York : Soc. Mot. Pict. Telev. Engrs).

Ferranti Ltd., Hollinwood, Lancs., Commercial Literature.

Fitzpatrick, J. A., Hubbard, J. C., and Thaler, W. J., 1950, *J. Appl. Phys.*, **21**, 1269.

Fraser, R. P., and Dombrowski, N., 1957, HSP3, p. 376.

Froome, K. D., 1948, *J. Sci. Instrum.*, **25**, 371 ; 1952, *Photogr. J.*, **92**B, 158.

Früngel, Dr. Ing. Frank, G.m.b.H., Hamburg, Germany, Commercial Literature ; Früngel, F., 1956, HSP2, p. 19 ; 1957, HSP3, p. 57.

Funfer, E., and Muller, W., 1956, HSP2, p. 244.

Geary, D. H., 1939, *Photogr. J.*, **79**B, 291.

General Electric Co. Ltd., Kingsway, London, Commercial Literature.

Germeshausen, K. J., 1949, *J. Soc. Mot. Pict. Telev. Engrs*, **52**, 24.

Gibson, F. C., Bowser, M. L., Ramaley, C. W., and Scott, F. H., 1954, *Rev. Sci. Instrum.*, **25**, 173.

Grimshaw, H. C., and Hardy, V. O., 1951, *Safety in Mines Res. Estab. Rep.*, No. 32.

Hadland, J., 1957, HSP3, Discussion at p. 158.

Haensel, H., and Schardin, H., 1956, HSP2, p. 315.

Hayes, R. A., 1955, *Industr. Lab.*, **6**, 6 ; 1956, *Industrial Image*, **1** (4), 9.

v. Heine-Geldern, R., 1956, HSP2, p. 238.

Henry, P. S. H., 1944, *J. Sci. Instrum.*, **21**, 135.

Hercock, R. J., 1947, *The Photographic Recording of Cathode Ray Tube Traces* (London : Ilford Ltd.).

Herden, R. B., 1957, *J. Photogr. Sci.*, **5**, 20.

Herzberg, G., and Walker, G. R., 1948, *Nature, Lond.*, **161**, 647.

Hills, H. F., 1957, HSP3, p. 279.

Holder, D. W., and North, R. J., 1956, *Optical Methods for examining the Flow in High-speed Wind Tunnels*, N.P.L./Aero/300 (London : National Physical Laboratory).

Jacobs, S. J., and Klebba, A. A., 1950, *Nav. Ordn. Lab. Memo.*, No. 10826.

Jenkins, J. A., and Chippendale, R. A., 1951, *J. Brit. Instn Radio Engrs*, **11**, 505.

Jones, E., 1928, *Proc. Roy. Soc.* A, **120**, 603.

Jones, G. A., 1952, *High Speed Photography* (London : Chapman and Hall).

Jones, G. A., and Eyles, E. D., 1949, *J. Soc. Mot. Pict. Telev. Engrs*, **53**, 502.

Judge, A. W., 1950, *Stereoscopic Photography* (London : Chapman and Hall), p. 297.

Kanolt, C. W., U.S.A. Pat. Nos. 1,150,374 ; 1,260,682 ; 2,158,660 ; Brit. Pat. No. 117342.

Karolus, A., 1956, HSP2, p. 78.

Laporte, M., 1956, HSP2, p. 1.

Lennard-Jones, J., and Mitchell, J. W., Brit. Pat. No. 574581.

Looms, J. S. T., and North, R. J., 1957, HSP3, p. 62.

Lunn, G. H., 1957, HSP3, p. 102.

Lunn, G. H., and Chippendale, R. A., 1957, *Electronic and Radio Engineer* (incorporating *Wireless Engr*), **34**, 156.

Luy, H., and Schade, R., 1956, HSP2, p. 59.

Marshak, I. S., 1957, HSP3, p. 30.

Meek, J. M., and Turnock, R. C., 1952, *Photogr. J.*, **92**B, 161.

Merlin, Gerin, Dubuit, Commercial Literature, and see Geradin, *Merger Magazine*, **18**, **37**.

Miller, C. D., 1949, *J. Soc. Mot. Pict. Telev. Engrs*, **53**, 479.

Miller, C. D., and Shaftan, K., 1952, *Prod. Engng*, **18**, 168.

Miller, C. D., and Scharf, A., 1954, *High Speed Photography*, **5**, 5 (New York : Soc. Mot. Pict. Telev. Engrs).

Muraour, H., and Michel-Lévy, A., 1934, *C.R. Acad. Sci., Paris*, **198**, 825, 1499, 1760, 2091 ; 1949, *Mémorial de l'Artillerie française*, **23**, 105, 867.

Naslin, P., and Vivie, J., 1956, HSP2.

O'Brien, B., and Milne, G., 1949, *J. Soc. Mot. Pict. Telev. Engrs*, **52**, 30.

Payman, W., Shepherd, W. C. F., and Woodhead, D. W., 1937, *Safety in Mines Res. Board Rep.*, **12**, 197.

Philips', N. V., Eindhoven, Netherlands, Commercial Literature.

Prime, H. A., and Saxe, R. F., 1949, *Proc. Instn Elect. Engrs*, Pt. II, **96**, 662.

Prudence, M. B., 1957, HSP3, p. 345.

Quinn, H. F., and Bourque, O. J., 1951, *Rev. Sci. Instrum.*, **22**, 101.

Quinn, H. F., McKay, W. B., and Bourque, O. J., 1950, *J. Appl. Phys.*, **21**, 995.

La Reliephotographie, Société pour l'Exploitation des Procédés de Photographie en Relief, Maurice Bonnet, (a) Brit. Pat. Nos. 638 551, 648 323, Brit. Prov. Pat. No. 24002/47 ; (b) Brit. Prov. Pat. Nos. 29577/45, 29578/45, French Pat. No. 745942 ; (c) Brit. Pat. Nos. 615 629, 619 948, 622 917, 638 551, Brit. Prov. Pat. No. 26923/45.

Richards, M. S., 1952, *Proc. Instn Elect. Engrs*, Pt. IIIA, **99**, 729, 757.

Saxe, R. F., 1957, HSP3, p. 126.

Saxe, R. F., and Chippendale, R. A., 1955, *Brit. J. Appl. Phys.*, **6**, 336.

Schardin, H., 1953, *Proceedings of the Royal Photographic Society Centenary Conference* (London : Royal Photographic Society), p. 388 ; 1954, *High Speed Photography*, **5**, 285 (New York : Soc. Mot. Pict. Telev. Engrs) ; 1956, HSP2, p. 301 ; 1957 a, HSP3, Discussion at pp. 114, 115 ; 1957 b, *Ibid.*, p. 316 ; 1957 c, *Ibid.*, p. 365.

Schardin, H., and Struth, W., 1937, *Z. Tech. Phys.*, **18**, 474.

Schultze, R. S., 1955, *Science and Applications of Photography* (*The Proceedings of the Royal Photographic Society Centenary Conference*, 1953) (London : Royal Photographic Society). Particularly : Chesterman, W. D., p. 356 ; Fraser, R. P., and Dombrowski, N., p. 360 ; Holder, D. W., and North, R. J., p. 371 ; Devaux, P., Fayolle, P., and Naslin, P., p. 377 ; Waddell, J. H., p. 384 ; Schardin, H., p. 388 ; Naslin, P., p. 393.

Scophony Ltd., 1946, *Photogr. J.*, **86**B, 42 ; 1947, Brit. Prov. Pat. No. 31752/47.

Smith, J. H., 1945, *J. Soc. Mot. Pict. Telev. Engrs*, **45**, 171.

Society of Motion Picture and Television Engineers, 1949–54, *High Speed Photography*, 1949, **1** ; 1949, **2** ; 1951, **3** ; 1952, **4** ; 1954, **5** (New York : Soc. Mot. Pict. Telev. Engrs).

Spencer, D. A., 1951, *Progress in Photography* 1940–50 (London and New York : Focal Press). Particularly : Jones, G. A., p. 191 ; Waddell, J. H., p. 200 ; Waldram, J. M., p. 323 ; 1954, *Progress in Photography*, 1951–54 (London and New York : Focal Press). Particularly : Bourne, H. K., p. 37 ; Shaftan, K., p. 201 ; Fage, A., p. 248.

Standring, W. G., and Looms, J. S. T., 1952, *Proc. Phys. Soc.* B, **65**, 108.

Stanton, J. S., and Blatt, M. D., 1948, The Bowen 76-lens Camera, *Nav. Ordn. Test Stn Rep.*, No. 152.

Suhara, T., 1934, *Proc. Imp. Acad. Japan*, **10**, 452.

Sultanoff, M., 1950 a, *Rev. Sci. Instrum.*, **21**, 653 ; 1950 b, *J. Soc. Mot. Pict. Telev. Engrs*, **55**, 158 ; 1954, *High Speed Photography*, **5**, 178 (New York : Soc. Mot. Pict. Telev. Engrs) ; 1956, HSP2, p. 230.

Tawil, E. P., 1957, HSP3, p. 9.

Tchernyi, J., 1957, HSP3, p. 324.

Thackeray, D. P. C., 1957, HSP3, p. 21.

Thompson, J. Langham, Ltd., 1957, *J. Sci. Instrum.*, **34**, 37.

Thun, R., 1956, HSP2, p. 132.

Turnock, R. C., 1951, *Proc. Instn Elect. Engrs*, Pt. II, **98**, 635.

Tuttle, F. E., 1949 a, *J. Soc. Mot. Pict. Telev. Engrs*, **53**, 451 ; 1949 b, *Ibid.*, **53**, 462.

Uyemura, T., 1957, HSP3, p. 300.

Uyemura, T., and Morishige, T., 1957, HSP3, p. 96.

Viard, J. L., 1957, HSP3, p. 108.

Vinten, W., Ltd., North Circular Road, London, Commercial Literature.

Waddell, J. H., 1946, *J. Soc. Mot. Pict. Telev. Engrs*, **46**, 87 ; 1954, *High Speed Photography*, **5**, 20 (New York : Soc. Mot. Pict. Telev. Engrs) ; 1955, *Photographic Motion Analysis* (Chicago : Industrial Laboratories Publishing Co.) ; 1957, HSP3, p. 351.

Walker, E. W., 1956, HSP2, pp. 91, 99 ; 1957, HSP3, p. 133.

Weibull, W., 1947, *Nature, Lond.*, **159**, 402.

Western Electric Co., New York City, N.Y., ' Fastax ' Commercial Literature.

Whitehead, J. R., and Fance, D. F., 1945, *Telecomm. Res. Estab. Rep.*, No. T2063.

Wilkinson, R. I., and Romig, H. G., 1954, *High Speed Photography*, **5**, 325 (New York : Soc. Mot. Pict. Telev. Engrs).

Winning, C. H., and Edgerton, H. E., 1952, *J. Soc. Mot. Pict. Telev. Engrs*, **59**, 178.

Wollensak, ' Fastax ' Commercial Literature.

Young, A. F. B., and Legg, D., 1955, *J. Photogr. Sci.*, **3**, 10.

Zarem, A. M., and Marshall, F. R., 1950, *Rev. Sci. Instrum.*, **21**, 514.

Zarem, A. M., Marshall, F. R., and Poole, F. L., 1949, *Elect. Engng*, **68**, 282.

Zavoiskii, E. K., and Fanchenko, S. D., 1955, *Doklady Akad. Nauk. S.S.S.R.*, **100** (4), 661.

Zernow, L., and Hauver, G. E., 1957, HSP3, p. 305.

PHOTOGRAPHIC SENSITIVITY

By J. W. MITCHELL
H. H. Wills Physical Laboratory, University of Bristol

CONTENTS

PAGE
§ 1. Introduction 435
§ 2. The physical properties of silver halide crystals 437
  2.1. Crystal imperfections 437
  2.2. Lattice defects and ionic conductivity 438
  2.3. Optical absorption 440
  2.4. Luminescence 440
  2.5. Electronic conductivity 440
  2.6. Crystal counters 441
  2.7. Displacement of the latent image and of photolytic silver in electric fields 442
  2.8. The diffusion of silver and gold 443
  2.9. Optical absorption due to products of photochemical change .. 443
  2.10. Photoconductivity associated with impurity absorption bands 444
  2.11. The photo-voltaic effect 444
§ 3. The experimental investigation of photographic sensitivity 444
  3.1. General background 444
  3.2. Silver halide systems employed for photographic research 445
§ 4. The primitive sensitivity of silver halide crystals 448
  4.1. General remarks 448
  4.2. The primitive sensitivity of photographic emulsions 448
  4.3. The primitive sensitivity of silver halide crystals 448
  4.4. Dislocations in silver halide microcrystals 449
  4.5. Sensitivity centres 451
§ 5. Chemical sensitization 453
  5.1. Introduction 453
  5.2. The sensitizers in normally processed active gelatins 453
  5.3. Methods of chemical sensitization 453
    5.3.1. Adjustment of silver and hydroxyl ion concentrations and digestion 453
    5.3.2. Reduction sensitization 454
    5.3.3. Sulphur sensitization 454
    5.3.4. Gold sensitization 455
  5.4. The nature of chemical sensitization 455
  5.5. Antifoggants and stabilizers 458
  5.6. Spectral sensitization arising from chemical sensitization 459
§ 6. Spectral sensitization with dyes 459
  6.1. Experimental observations on dye sensitization 459
  6.2. The theory of dye sensitization 462
§ 7. The formation and properties of the latent image 464
  7.1. General observations 464
    7.1.1. Theories of the formation of the latent image 464
    7.1.2. Intermediate steps in the formation of latent image specks 465
    7.1.3. Observations on the theory of the formation of latent image specks 466
  7.2. Energy levels in crystals of silver halides 467
    7.2.1. Energy levels within the crystals 467
    7.2.2. The absorption of energy within the crystals 467

PAGE
7.2.3. Surface energy levels 469
7.2.4. Surface energy levels due to adsorbed chemical sensitizers 469
7.2.5. Optical transitions involving levels in the forbidden band 470
7.2.6. The sensitivity centres 471
7.3. The primary stage in the formation of the latent image 471
7.4. The secondary stage in the formation of the latent image 473
7.4.1. Latent pre-image specks 473
7.4.2. Surface and internal latent sub-image specks 475
7.4.3. Chemical sensitization and the formation of the latent sub-image 475
7.4.4. The formation and properties of latent image specks 476
7.5. Conclusion 478
§ 8. Reciprocity failure 479
8.1. Introduction 479
8.2. The interpretation of reciprocity failure 481
8.3. The reduction of reciprocity failure 483
§ 9. Latent image effects 484
9.1. The formation of the latent image at low temperatures 484
9.2. The Herschel and Debot effects 484
§ 10. Solarization and photolysis 486
10.1. Solarization 486
10.2. The photolysis of silver halides 487
10.2.1. Introduction 487
10.2.2. Experimental work on photolysis in emulsions 488
10.2.3. Experimental work on photolysis in large crystals 488
(*a*) The quantum yield for the production of halogen 488
(*b*) The quantum yield for the production of silver 489
10.3. The orientation of the silver produced by electron bombardment and photolysis with respect to the silver halide 490
10.3.1. General remarks 490
10.3.2. Experimental observations 491
10.4. The theory of photolysis 491
10.5. The principles underlying the sensitization of ionic crystals for photochemical change 493
§ 11. Development 495
11.1. The surface and the internal latent image 495
11.2. The kinetics of development 495
11.2.1. General remarks 495
11.2.2. Factors which influence the induction period 495
11.2.3. Latensification 496
11.2.4. Development accelerators 497
11.3. The form of the silver produced by development 498
11.4. Theories of development 499
11.4.1. The triple-interface theory 499
11.4.2. The Gurney–Mott theory of development 500
11.4.3. A revised electrode theory of development 501
11.5. Conclusion 503
§ 12. General conclusions 504
References 508
Reports of photographic conferences 508
Useful reference works 508
Original papers 508

Abstract. This paper reviews the experimental and theoretical work on photographic sensitivity and related topics which has been published during the past ten years. New experimental results, obtained to a large extent with model systems, have necessitated substantial revision and amplification of previously accepted theories. The role of crystal imperfections in all aspects of photographic sensitivity is now recognized and emphasis has shifted from the trapping of electrons, which characterized the Gurney–Mott theory, to the trapping of positive holes. The latent image consists of small groups of silver atoms formed by the combination of silver ions and electrons. This proceeds through intermediate stages in which first latent pre-image and then latent sub-image specks appear. The initiation of development by latent image specks of different sizes is discussed in the final sections. The approach which is followed in this paper leads to an integrated treatment of the whole range of photographic phenomena in which very few hypotheses with no direct experimental foundation have to be introduced.

---

§ 1. INTRODUCTION

TEN years have passed since the last review of the subject of photographic sensitivity in these reports. During these years, intensive research activity has been stimulated by a series of colloquia and conferences, the published proceedings of which provide a record of substantial progress (Liège 1948, edited by Hautot 1948 ; Zürich 1949, edited by Eggert and Ammann 1950 ; Bristol 1950, edited by Mitchell 1951 ; Paris 1951, edited by Morand and Vassy 1953 ; Hakone 1953, edited by Fujisawa 1956 ; London 1953, edited by Schultze 1955 ; and Cologne 1956, edited by Frieser 1957).

In 1947, it appeared that the Gurney–Mott theory of the formation of the latent image had been firmly established by the results of experimental work carried out principally in the Kodak Research Laboratories at Rochester and Harrow. Sheppard (1928) had concluded that a distribution of discrete specks of silver sulphide was produced, during chemical sensitization, on the surfaces of silver halide microcrystals in emulsions. He had suggested that the function of these specks was to concentrate the silver atoms which are liberated by the absorption of energy in the crystals (Sheppard, Trivelli and Loveland 1925). Gurney and Mott (1938) proposed that electrons, released when energy was absorbed in the crystals, were trapped by the silver sulphide specks, or by particles of silver, to charge them negatively. The negative charge was then neutralized by the motion of silver ions to the specks, so that silver atoms were formed on them. This mechanism was first introduced to explain why discrete particles of silver appear during photolysis, and then extended to the formation of the surface latent image which requires the absorption of considerably smaller amounts of energy in the crystals. Berg (1948) gave a full account of the application of the Gurney–Mott theory to the interpretation of the different latent image phenomena.

Two other aspects of photographic sensitivity were not considered in detail either by Sheppard or by Gurney and Mott :

(1) The possible role of crystal imperfections (Trivelli 1909, 1923, Steigmann 1924, 1930, and Smekal 1932).

(2) The possible reaction between the chemical sensitizer or the silver particle, and liberated halogen (Hickman 1927).

Berg (1948) did not make any reference to the above publications, which shows that little emphasis was then placed on these features of photographic sensitivity.

The concentration speck theory and the Gurney–Mott theory in their original forms were concerned mainly with the formation of the surface latent image. In neither theory was any particular role attributed to crystal imperfections and it was assumed that the halogen was simply liberated at the surfaces of the crystals where it reacted with gelatin (Sheppard and Graham 1940, Mott 1941).

Considerable stress has been placed on both these aspects of photographic sensitivity in the work at Bristol since 1947. It has been shown that the internal latent image is formed during exposure, by the separation of silver atoms along dislocation lines (Hedges and Mitchell 1953 a, b, Evans, Hedges and Mitchell 1955, Clark and Mitchell 1956) and that silver sulphide is formed preferentially where dislocations meet the surface of the crystal (Mitchell 1953 b, 1954, 1957 a, b). The products of chemical sensitization are found to have a number of functions :

(1) They extend the optical absorption and photographic sensitivity of the crystals to longer wavelengths (Eggert and Biltz 1938, 1939, 1940, Eggert and Kleinschrod 1939, 1940 a, b).

(2) They trap positive holes or react with halogen (Hickman 1927, Burrow and Mitchell 1954, Mitchell 1956, 1957 a, b, d).

(3) They provide centres for the combination of silver ions and electrons to form silver atoms (Eggert 1926, Eggert and Rahts 1928).

(4) They hinder the diffusion of silver atoms away from the surfaces of the crystals (Evans, Hedges and Mitchell 1955).

The Bristol work has led to a theory of photographic sensitivity which will now be briefly summarized in order that the significance of the different sections of this report may be fully appreciated. Positive holes and electrons are liberated when energy is absorbed within silver halide crystals. The positive holes are then trapped by surface halide ions or by adsorbed silver sulphide molecules and their positive charge appears as an equivalent charge due to excess silver ions which is compensated by the negative charge of the conduction electrons. During the interval between their liberation and their combination with silver ions, the electrons may be displaced in electric fields so that photoconductivity (Toy and Harrison 1930, West 1951), the displacement of the latent image (Hedges and Mitchell 1953 b, Evans, Hedges and Mitchell 1955, Hamilton, Hamm and Brady 1956) and of photolytic silver (Haynes and Shockley 1948, 1951, Webb 1955) may be observed.

Latent image specks are formed by the combination of silver ions and conduction electrons which proceeds through a series of intermediate stages. In the first, latent pre-image specks (Mitchell 1957 a, b) are produced by the combination of interstitial silver ions and electrons at sites adjacent to surface silver *ions*.† These specks consist of atoms of silver, adsorbed on the external and internal surfaces of the crystals ; they have a short lifetime at any temperature at which interstitial silver ions are mobile. During this lifetime, the specks may combine with a further interstitial silver ion and conduction electron to form more stable latent sub-image specks. These are therefore pairs of silver atoms adsorbed at sites adjacent to surface silver *ions* ; they do not furnish effective

† See § 7.3.

traps for conduction electrons at room temperature but do trap positive holes and may therefore suffer regression. The latent sub-image specks are transformed into stable latent image specks of minimum size by combination with an interstitial silver ion and conduction electron. At this stage, a critical change in the mechanism takes place: the group of three silver atoms, which has been formed from interstitial silver ions and electrons by the processes which have been outlined, can adsorb a silver ion from the surface of the silver halide crystal and thus become positively charged in thermal equilibrium at room temperature, the compensating negative charge being provided by a vacant silver ion lattice site in the neighbourhood. These and all larger groups of silver atoms which are positively charged repel positive holes, and are therefore protected from regression. They trap conduction electrons and the positive charge is then immediately restored to the uncharged speck by the adsorption of a further silver ion from the surface and the formation of a vacant silver ion lattice site. The potential differences arising from the trapping of the positive holes and electrons at separated sites are eliminated by the motion of silver ions and, through a succession of such processes, the silver specks increase in size until they become visible particles of photolytic silver.

Photolytic silver is thus formed in discrete particles by a two-stage mechanism, as was postulated by Gurney and Mott (1938): there is, however, the fundamentally important difference that the adsorption of a silver ion to the particle precedes the trapping of the electron. It became evident during the course of the experimental work at Bristol that the theory of photographic sensitivity, as it stood in 1947 (see Berg 1948), required substantial revision. The new theory (Mitchell 1953 b, 1954, 1956, 1957 a, b) is presented here for the first time in its relationship to the contributions made by many other workers in different countries during the same period.

§ 2. THE PHYSICAL PROPERTIES OF SILVER HALIDE CRYSTALS

2.1. *Crystal Imperfections*

Crystal imperfections play an important role in the photographic process (Mitchell 1952, 1954, 1957 a, b). That this might be the case was suggested by Smekal (1932) on the basis of observations made by Rexer (1931, 1932 a, b) on imperfections in crystals of sodium chloride. The first actual observations of well-defined imperfections in silver bromide crystals were made by Hedges and Mitchell (1953 a), who found that they could render sub-structures visible by printing silver out along the individual dislocation lines.

Since then, sub-structures of silver chloride crystals and of silver chlorobromide and silver bromo-iodide mixed crystals have been examined. The motion of the dislocations during plastic deformation has been studied and many observations made which are relevant to the discussion of photographic problems such as kink desensitization. The enhanced reactivity of the areas around the points in which the dislocations meet the external surfaces has been demonstrated and made the basis of methods for locating them. Specks of silver are produced at these sites when crystals are exposed to long-wavelength ultra-violet light under solutions of a number of halogen acceptors (as, for example, 1-phenyl-3-pyrazolidinone), and etch pits are formed at the same sites when the crystals are treated with solutions of silver halide solvents (Evans and Mitchell 1955, Mitchell 1957 c).

2.2. *Lattice Defects and Ionic Conductivity*

In the range of temperatures in which measurements of ionic conductivity may conveniently be made, the cation transport number for crystals of silver halides is unity (Tubandt 1932). The lattice of cations is disordered, and the lattice defects are of the Frenkel type (Jost 1933, 1938, Wagner and Beyer 1936). In pure crystals they consist of pairs of interstitial silver ions and vacant silver ion lattice sites.

The work of Koch and Wagner (1937) on the influence of divalent cations, such as lead and cadmium ions, on the ionic conductivity of large crystals of silver halides has been extended by Teltow (1949 a, b) and by Ebert and Teltow (1955). The substitution of divalent cations for silver ions on lattice sites in increasing numbers first reduces the ionic conductivity to a minimum value which is less than the conductivity of the pure crystal, and then increases it to much greater values. The initial decrease arises from the suppression of the interstitial silver ions, and the subsequent increase from the introduction of vacant silver ion lattice sites to compensate the excess positive charges of the divalent cations. Ebert and Teltow (1955) have derived the values shown in table 1 for the formation energies of pairs of Frenkel defects and the activation energies for the diffusion of the individual defects (the small correction which may be made on account of the association of divalent cations and vacant silver ion lattice sites has *not* been included).

Table 1. Energies (in ev) of Lattice Defects

	Formation energy of Frenkel defect	Activation energy for migration of	
		Interstitial silver ion	Vacant silver ion lattice site
Silver chloride	1·69	0·15	0·33
Silver bromide	1·27	0·15	0·36

All the available experimental results show that silver ions require a smaller activation energy for diffusion by way of interstitial positions than by any other mechanism. The high mobility of the interstitial silver ions which has been established is thus of great photographic interest. The work of Teltow shows that there are about 10^{12} Frenkel defects per cm^3 in a pure silver bromide crystal at room temperature. Teltow has also studied the effect of monovalent ions, including cuprous and iodide ions (1950 a), and of dissolved silver, cadmium and lead sulphides (1950 b) on the ionic conductivity of silver bromide crystals. He concluded that cuprous bromide dissolved in silver bromide largely in the form of interstitial cuprous ions and vacant silver ion lattice sites, and that silver sulphide dissolved as divalent sulphide ions associated with interstitial silver ions. A similar conclusion has been reached by Ebert and Teltow (1955) from their work on the corresponding silver chloride mixed crystals.

The occurrence of polarization effects in measurements of the ionic conductivity of silver bromide crystals has been described by Friauf (1954). Jost, Mennenöh and Müller (1949), Jost and Mennenöh (1950) and Mennenöh (1950) have continued the earlier work of Jost and Nehlep (1936 a, b) on the effect of pressure on the ionic conductivity of crystals of silver halides, and similar work has also been undertaken by Kurnick (1952). The results are in general agreement with the conclusion that the lattice disorder is of the Frenkel type.

The hypothesis introduced by Mitchell (1948, 1949 a, b, 1951 a, b, see also Anastasevich and Frenkel 1941), that Schottky defects consisting of pairs of vacant silver and vacant halide ion lattice sites might be involved in the formation of the internal latent image and in chemical sensitization at room temperature, has not been supported by subsequent investigations (Berry 1951, 1955, Keith and Mitchell 1951, Mitchell 1957 d) and has been discarded. The need for this hypothesis disappeared when it was demonstrated that the internal latent image was formed by the separation of silver atoms along dislocation lines. Stasiw (1950, 1951, 1952 a, b, c, 1954) has postulated that vacant halide ion lattice sites form association complexes with divalent anionic impurities in silver halide crystals which have been quenched from temperatures near the melting point. This will be discussed in §§ 5.4 and 10.2.3 (*b*). Schottky defects may also be present at temperatures near the melting point in pure silver bromide crystals (Kanzaki 1951, 1956, Kurnick 1952) and their condensation at lower temperatures may introduce dislocation networks by the mechanisms proposed by Seitz (1950 c). Seitz (1950 a, b, d) has discussed the influence of the generation of lattice defects during plastic deformation on the electrical and photochemical properties of alkali halide crystals. The effect of plastic deformation on the electrical conductivity of silver bromide crystals has been measured by Johnston (1955).

The drift of silver ions over the surfaces of the particles of lightly compacted silver bromide powders was observed by Shapiro and Kolthoff (1947), who deduced the following equation for the surface ionic conductivity:

$$\sigma = \text{constant} \times S \times \exp(-0{\cdot}36/\boldsymbol{k}T),$$

where S is the surface per unit mass of the powder and 0·36 ev is an activation energy. The drift of silver ions along grain boundaries at a higher rate than through the crystal lattice was demonstrated in transport experiments made by Goddard and Urbach (1952). In view of current ideas on the properties of dislocations, it would be surprising if silver ions did not diffuse along edge dislocations.

So far, we have considered the properties of large silver halide crystals and of dry compacted powders (Shapiro and Kolthoff 1947). The properties of the microcrystals suspended in aqueous solutions are also of interest in photographic sensitivity. The modification of the lattice defect concentrations which occurs near the interface between silver halide crystals and aqueous solutions was investigated by Grimley and Mott (1947, see also Grimley 1950 a, b, 1951). Holliday and Davies (1951) and Davies and Holliday (1952 a, b) observed the effect of silver ion concentration on the electrophoretic mobility of colloidal particles of silver bromide. They adjusted the silver ion concentration with silver fluoride instead of with silver nitrate, as had usually been done previously. Zimen (1946, 1951) studied the exchange of radioactive silver and bromide ions between particles of silver bromide and the surrounding solution, and found that the exchange of silver ions occurred rapidly at a rate consistent with the known mobilities of Frenkel defects, whereas the exchange of bromide ions occurred at an extremely small rate. This confirms the conclusion that vacant bromide ion lattice sites cannot play any important role in chemical or photochemical processes in pure silver bromide crystals at room temperature. The theory of the radioactive ion exchange experiments has been examined by Pitts (1954).

2.3. *Optical Absorption*

Following the earlier work of Slade and Toy (1920) and Hilsch and Pohl (1930), the optical absorption of silver halide crystals near the absorption edge has been measured by Moser and Urbach (1956). The interpretation of the temperature dependence of the absorption edge has been discussed by Urbach (1953). The measurements on silver bromide have since been extended to much shorter wavelengths and lower temperatures by Okamoto (1956), who used thin films deposited under good vacuum conditions and found a series of exciton bands on the short-wave side of the absorption edge. Raynaud and Pouradier (1955 b) have measured the reflection spectra of silver bromo-iodide precipitates and confirmed the extension of the absorption to wavelengths beyond the absorption edge of pure silver bromide precipitates, which had been deduced by Huse and Meulendyke (1926) from the wavelength dependence of the photographic sensitivity of silver bromo-iodide emulsions.

The theory of the ultra-violet absorption of silver bromide crystals has been discussed by Seitz (1951, 1956) and by Allen (1953). The hypothesis that the characteristics of the absorption band near the long-wave limit arise from transitions at crystal imperfections which was advanced by Seitz has not been supported by the results of experimental investigations.

2.4. *Luminescence*

The silver halides become luminescent when they are cooled to low temperatures and this phenomenon is of considerable interest because it may determine the quantum efficiency with which the latent image is formed at low temperatures (Farnell 1952, Biltz 1952).

The luminescence was noted by Randall (1939) and was studied by Meidinger (1940, 1944), Farnell, Burton and Hallama (1950) and by Farnell and Burton (1951), in both silver halide precipitates and photographic emulsions.

2.5. *Electronic Conductivity*

The photoconductivity of crystals of silver bromide was first studied quantitatively at low temperatures by Hecht (1932) and Lehfeldt (1935), who found no evidence for the motion of positive holes. At room temperature, the electrical conductivity may decrease or increase on illumination (Toy and Harrison 1930, West 1951), and the behaviour of the crystals is complicated by the relatively large ionic dark conductivity. West attributed the decrease in conductivity to the combination of mobile silver ions with electrons to form immobile silver atoms. Yamada and Oka (1956) have, however, shown that this effect disappears when the crystals are treated with oxidizing agents, the anticipated increase in conductivity on illumination then being observed. It thus appears that the mobile silver ions do not combine with electrons unless nuclei which can be destroyed by treatment with oxidizing agents are available. Grenishin and Minkov (1954) have shown that the photocurrents at room temperature are increased to a maximum by the incorporation of between 0·1–0·5 mol per cent of cadmium bromide in the silver bromide.

The properties of conduction electrons in crystals of silver halides at room temperature are best established by releasing the electrons with sharply defined light pulses and displacing them by the application of synchronized voltage pulses. This technique was introduced by Haynes and Shockley (1948, 1951) and their measurements on silver chloride crystals are summarized in table 2.

Table 2. Properties of Electrons in Silver Chloride Crystals at Room Temperature

Drift mobility μ_D = 49·5 cm^2 v^{-1} sec^{-1}
Hall effect mobility μ_H = 51 ± 3 cm^2 v^{-1} sec^{-1}
Mean free lifetime = 10^{-6} sec.

The closeness of the values for the mobility derived from observations of drift in electric fields and from the Hall effect shows that the electrons spend a negligible fraction of their lifetime in shallow traps at room temperature. Suptitz (1957) has recently made similar measurements with large crystals of silver bromide. Mobilities of positive holes in crystals of silver halides at room temperature have not yet been determined with this experimental technique.

The diffusion of positive holes in crystals of silver halides at room temperature has been established by Stasiw and Teltow (1944), Teltow (1951, 1952 a) and Jung, Stasiw and Teltow (1951), who have observed the bleaching of the brown coloration of crystals of silver chloride and silver bromide containing traces of silver sulphide which occurs when the crystals are exposed to halogens. A more direct demonstration of the mobility of positive holes at room temperature in crystals of pure silver bromide has recently been provided by Luckey and West (1956), who have observed a rapid increase in the conductivity of silver bromide crystals on exposure to halogen and an equally rapid decrease upon the removal of the halogen by evacuation. From their observations, it may be deduced that the mobility of positive holes is greater than that of interstitial silver ions and that positive holes are not trapped by vacant silver ion lattice sites at room temperature.

2.6. *Crystal Counters*

The observations which have been made on the current pulses which are produced when β and γ rays and nuclear particles are absorbed in crystals of silver bromide are of interest both from the point of view of the formation of the latent image in nuclear emulsions and from the information which they provide on the mobilities and lifetimes of conduction electrons and positive holes. Van Heerden (1945, 1950) first observed that current pulses were produced when single ionizing particles passed through crystals of silver chloride in electric fields of the order of 1000 v cm^{-1} at the temperature of liquid nitrogen. With β particles, the loss of energy was estimated to be 7·6 ev for each electron and positive hole pair created (van Heerden and Milatz 1950). Yamakawa (1951) then investigated the properties of silver bromide crystal counters and found a value of 5·8 ev for the energy required to produce an electron and positive hole pair. He deduced that the mobility of the electrons was 210 cm^2 v^{-1} sec^{-1} and that their mean free path was 5×10^{-4} cm^2 v^{-1} at 77°K. Similar measurements have been made between 77°K and 200°K by Allemand and Rossel (1954 a, b) on both silver chloride and silver bromide crystals. Their results are summarized in table 3.

Table 3. Properties of Electrons in Silver Halide Crystals at 77°K

	Silver chloride	Silver bromide†
Drift mobility μ_D (cm^2 v^{-1} sec^{-1})	240 ± 20	72 ± 8 62 ± 10
Mean free lifetime (sec)	$6{\cdot}3 \times 10^{-6}$	$3{\cdot}6 \times 10^{-6}$ $1{\cdot}2 \times 10^{-6}$
Density of traps (cm^{-3})	$3{\cdot}0 \times 10^{13}$	$1{\cdot}5 \times 10^{14}$ $5{\cdot}3 \times 10^{13}$

† Results obtained with two crystals.

Brown (1953) found a value for the Hall mobility of electrons in silver chloride crystals of 370 ± 70 cm^2 v^{-1} sec^{-1} at 89 ± 6°K. He has since measured the drift mobility of electrons and found a value of 274 cm^2 v^{-1} sec^{-1} at 86°K (Brown 1955). In all this work on the current pulses produced by the passage of β rays through silver halide crystals at low temperatures, no pulses which could be specifically attributed to the displacement of positive holes have been observed.

The induced conductivity arising from the bombardment of silver chloride crystals with electron pulses with energies up to 10 000 ev has been studied by Moore (see Seitz 1951 and Smith 1951) and by Yamanaka and Suita (1955), both of whom have observed current pulses due apparently to the displacement of positive holes.

2.7. *Displacement of the Latent Image and of Photolytic Silver in Electric Fields*

The displacement of the surface and internal latent images away from the illuminated area of large crystals of silver bromide in electric fields was first observed by Hedges and Mitchell (1953 b) and Evans, Hedges and Mitchell (1955). The displacement of the latent image in microcrystals of photographic emulsions has been observed by Hamilton, Hamm and Brady (1956), who have also obtained evidence for the displacement of positive holes. The crystals are illuminated in these experiments with light flashes of short duration and the displacement is produced by the application of synchronized voltage pulses as in the experiments of Haynes and Shockley (1948, 1951), whose results have already been reported in §2.5. The method has been applied in a particularly elegant form by Webb (1955) to demonstrate the displacement of photolytic silver in large silver bromide crystals and in microcrystals from emulsions. From his observations, Webb deduced values of 50 cm^2 v^{-1} sec^{-1} for the mobility and 10^{-5} sec for the lifetime of the electrons.

All these observations show that the separation of silver in crystals of silver halides involves the liberation, displacement and trapping of conduction electrons.

Potential differences arise from the trapping of the electrons and positive holes at separated sites. The silver halides are ionic conductors at room temperature, so that silver ions drift in the electric field to neutralize the space charges. These observations thus establish the validity of the Gurney–Mott principle that the motion of electrons and the motion of silver ions is involved in the formation of the latent image and of photolytic silver in crystals of silver halides, and not just that of electrons, as had previously been assumed (Gurney and Mott 1938).

2.8. *The Diffusion of Silver and Gold*

Evans, Hedges and Mitchell (1955) have shown that silver is able to diffuse over the external surfaces and into the interior of crystals of silver bromide at room temperature. They deposited sharply defined patches of silver by vacuum evaporation through masks and established the existence of surface diffusion by treating the crystals with physical developers before and after periods of storage. The occurrence of diffusion along the dislocation lines of the sub-boundaries was demonstrated by physical or chemical development after the removal of the surface layers by treatment with a dilute solution of potassium cyanide and oxidation of residual surface silver. No evidence for the diffusion of gold was found when the same experiments were repeated with thin films of gold.

2.9. *Optical Absorption due to Products of Photochemical Change*

The absorption of energy in crystals of silver halides at room temperature causes the liberation of halogen and the separation of silver on internal imperfections. No absorption which can be attributed to isolated silver atoms has ever been observed to arise from the irradiation of crystals of pure silver halides (Hilsch and Pohl 1930, 1932). Isolated silver atoms must have a very short lifetime at any temperature at which silver can separate so that the silver ions and electrons condense to larger aggregates. The resultant ' latent image ' absorption band was studied by Hilsch and Pohl (1930). The selective bleaching which they observed when the crystals were subsequently exposed to monochromatic light showed that particles of different sizes contributed to this absorption. Kaiser (1952) has more recently condensed silver atoms and silver chloride or silver bromide molecules upon transparent substrates cooled to low temperatures and found an absorption band which may be attributed to isolated silver atoms adsorbed upon internal surfaces. This band disappears, however, when the composite films are warmed up to room temperature and is replaced by the broad ' latent image ' absorption band.

A change in optical absorption due to products of photochemical change formed at low temperatures in crystals of silver halides containing traces of silver sulphide or silver selenide has been found by Stasiw and his co-workers. Their results will be presented in § 10.2.3 (*b*).

Observations of a change in optical absorption accompanying the formation of a latent image were first made by van Kreveld and Jurriens (1937). These measurements have now been repeated with an improved spectrophotometer by Nail, Moser and Urbach (1956), who used a wavelength of 6000 Å for measurements of the change in optical density accompanying the formation of a latent image in photographic emulsions. The use of radioactive iodine for determining the silver in photographic images has been described by Ballard, Stevens and Zuehlke (1952).

2.10. *Photoconductivity associated with Impurity Absorption Bands*

Meiklyar and Putseiko (1950) have observed a number of bands of sensitized photoconductivity beyond the edge of the lattice absorption band in illuminated crystals of silver chloride and silver bromide. Similar observations have been made by Gilleo (1953 a, b), who has observed bands of sensitized photoconductivity at 90°K with crystals of silver chloride covered with thin vacuum-deposited films of silver. These bands have maxima at 4400 Å, 5000 Å, 5700 Å, 6600 Å and 7500 Å. Suzuki (1956) has observed sensitized photoconductivity in crystals of silver bromide containing traces of silver sulphide, on exposure to infra-red radiation. All these observations suggest that the impurities, groups of silver atoms or silver sulphide molecules, introduce occupied levels into the forbidden bands of the silver halide crystals from which electrons may be displaced to the conduction band by the absorption of light quanta. A consequence of this is that the same impurities should provide effective traps for positive holes in the full band of the crystals.

2.11. *The Photo-voltaic Effect*

The liberation and diffusion of electrons during the illumination of silver bromide surfaces at room temperature is clearly demonstrated by the photovoltaic effect which has been investigated with silver–silver bromide electrodes by Vanselow and Sheppard (1929), Sheppard, Vanselow and Hall (1929) and Sheppard, Vanselow and Happ (1940). Amsler (1953) has now described the preparation and properties of silver bromide photocells in which a layer of silver bromide is formed on a silver plate and covered with a thin transparent film of silver by evaporation. His work is of particular interest on account of the close correspondence which he observed between the wavelength dependence of the photocurrents in silver bromide photocells sensitized with silver, the absorption spectra of chemically sensitized photographic emulsions, and the photographic sensitivity of the emulsions (§ 5.6). It establishes that the absorption of light in the impurity absorption band associated with excess silver liberates conduction electrons. The cells may be sensitized by the adsorption of a layer of a suitable dye on the silver bromide surface before the deposition of the transparent silver film. Photocurrents are then observed when the cells are illuminated with light which is absorbed by the dye (§ 6.1).

§ 3. THE EXPERIMENTAL INVESTIGATION OF PHOTOGRAPHIC SENSITIVITY

3.1. *General Background*

A commercial silver bromo-iodide photographic emulsion presents an exceedingly complex system in which microcrystals with a wide range of shapes, sizes and compositions are suspended in gelatin or in a synthetic medium such as polyvinyl alcohol. The behaviour of the individual crystals is determined by their state of imperfection and by the nature of the substances which participate in the formation of adsorbed layers upon their surfaces by dynamic and competitive processes during sensitization. The statistical behaviour of the whole assembly of microcrystals may be investigated by the standard methods of photographic sensitometry, and many valuable conclusions drawn from the observations.

It is, however, necessary to analyse the behaviour of the system in greater detail, and the approach of the physicist to the elucidation of the mechanisms of chemical and photochemical change differs from that of the chemist. The physicist, following his training and natural inclination, tends to focus his attention upon the properties of the individual microcrystals in the emulsion which may be established by observations with the optical and electron microscopes and to supplement these observations with measurements of the physical properties of large silver halide crystals such as have been summarized in the previous section. The chemist is usually more concerned with the surface phenomena involved in chemical and optical sensitization and in exposure, and with the application of all the chemical and physical methods of analytical chemistry to their investigation.

For neither approach does a commercial photographic emulsion provide a suitable system. Although much progress has been made in recent years with the application of the techniques of electron microscopy (Hamm and Comer 1953, Bradley 1954), the grain size of the average emulsion is usually too small to permit the investigation of the properties of the individual crystals by physical methods. The grain size of the average emulsion is too large for the application of the usual methods of analytical chemistry to the study of the chemical and photochemical changes which occur at the interfaces, because insufficient products are formed on account of the small specific surface area.

Because of these difficulties, there has been a rapid increase in recent years in the number of papers published on the results of investigations with model photographic systems which are more amenable to detailed study than the photographic emulsion itself.

3.2. *Silver Halide Systems employed for Photographic Research*

The silver halides are being employed in a remarkably wide range of forms for current researches on the nature of photographic sensitivity. This range extends from silver halide hydrosols, in which the tiny particles may be observed, under the intense illumination of the ultra-microscope, as diffraction discs in rapid Brownian motion, to large single crystals grown slowly from the molten substance and then sectioned and polished. From these two extremes, the normal photographic emulsion may be approached through a series of intermediate stages in which the crystal sizes are increased or reduced and gelatin is introduced as a protective colloidal medium. We may classify the systems which are being employed as follows :

(1) Silver halide hydrosols.

(2) Ultra-fine-grain emulsions prepared by adding gelatin or a synthetic medium to the hydrosols which are then exposed and processed in liquid form.

(3) Fine-grain silver halide emulsions, exposed and processed in liquid form.

(4) Slow fine-grain commercial positive photographic materials.

(5) High-speed commercial negative photographic materials.

(6) Coarse-grained pure silver bromide emulsions in which tabular microcrystals, 1 to 10 μ in diameter with apparent (111) surfaces, are produced by the slow addition of silver nitrate to solutions of alkali bromides in gelatin.

(7) Tabular microcrystals, 1 to 60 μ in diameter, grown by isothermal dilution from colloid-free aqueous solutions of silver halides in potassium bromide, hydrogen bromide, ammonium hydroxide, or other suitable solvents.

(8) Large thin single crystal sheets of fused silver halides, 1 to 100 μ in thickness, with (001) or (111) surfaces, cut into sections 6 mm square and mounted on 3 in. × 1 in. microscopic slides.

(9) Polished slabs, cut from large single crystals of silver halides.

Following the work of Lottermoser (1907, 1908), hydrosols were first used by Fajans and his co-workers (Fajans and Frankenburger 1923, Fajans and Steiner 1927, Fajans and Karagunis 1929) for the study of the influence of different variables, particularly silver ion and hydroxyl ion concentrations on the photolytic sensitivity of the silver halides. Similar experiments were continued by Urbach and Urbach (1950, see also Urbach 1935) who observed the fading of photolytic silver in silver bromide hydrosols and noted that this was prevented by the addition of gelatin but not by halogen acceptors. The first work on the formation and development of a latent image in silver halide hydrosols has been undertaken by Loening in the Kodak Research Laboratory at Harrow. Loening, who has worked out sound and reproducible techniques for the preparation, exposure and processing of hydrosols, has been able to study many features of the photographic system and, by adding gelatin to the hydrosol, to produce an exceedingly fine grained photographic emulsion which may be exposed and processed either in the liquid or the coated form. The addition of gelatin stabilizes the surface latent image which otherwise fades rapidly on storage. The internal latent image does not fade either in the presence or in the absence of gelatin (Loening 1951 a, b).

Loening has also studied the chemical sensitization of silver bromide hydrosols by the adjustment of the silver ion concentration (1952) and by the addition of soluble sulphides (1953). Bourne and Loening (1953) have considered the photographic implications of the adsorption by hydroxyl ions to silver bromide on the basis of their experiments with hydrosols.

The silver halide hydrosols and fine grained liquid emulsions are particularly suitable for research on the influence of different physico-chemical variables on the behaviour of particles of silver halides by sensitometric or by chemical methods. A uniformity of exposure and of processing is possible which cannot be achieved in coated emulsions, and the very great specific surface area of the hydrosol allows analytical methods to be used for the study of surface reactions (Sutherns and Loening 1956 a, b). The main disadvantage of the hydrosol for work on the formation and properties of the latent image is that the reaction mechanisms and the behaviour of the latent image may be modified in the liquid compared with the dry system. The removal of unwanted by-products from sensitizing chemical reactions and processing procedures also presents difficulties. The particles in the sols, even when examined with the electron microscope, are found to have no well-defined crystallographic forms, and the microscopic study of their behaviour is scarcely possible.

Fortmiller and James (1951, 1952, 1953) have exposed and processed commercial fine grained positive emulsions in liquid form. Coated fine grained positive emulsions have also been used by Lowe, Jones and Roberts (1951) for work on primitive sensitivity (see § 4) and by James and his co-workers for many kinetic studies on photographic development (see § 11). Coarse grained pure silver bromide emulsions, prepared by the methods described by Trivelli and Smith (1939, 1940), have been used by Evans and Mitchell (1955) for the study of dislocations in microcrystals (§ 4.4), by Webb (1955) for experimental work on the displacement of photolytic silver in electric fields, and by Hamilton, Hamm and Brady (1956) for work on the displacement of the latent image under similar conditions (§ 2.7). Large silver halide microcrystals prepared in the absence of gelatin have been used by Evans and Mitchell (1955), by Sutherns (1956) and by Fujisawa, Mizuki and Kubotera (1957), for investigations of crystal imperfections, photographic sensitivity and chemical development respectively.

Single crystals in the form of large plane-parallel sheets with optically flat surfaces were first used extensively by Mitchell and his co-workers for researches on photographic sensitivity. Full details of their preparation are given in the paper by Clark and Mitchell (1956). They have also been used by Luckey (1953, 1955) for the quantitative study of vacuum photolysis (see § 10.2.3 (*a*)), by Luckey and West (1956) for experiments on the diffusion of positive holes at room temperature (see § 2.5), and by West and Saunders (1957) for experiments on the formation of the internal latent image and the adsorption of sensitizing dyestuffs (see § 6.1). Polished surfaces of large single crystal slabs and spheres of silver chloride and silver bromide have been used by Boissonnas (1949 a, b, 1950, 1951), Vuille (1951, 1954 a, b) and by Leidheiser and Cook (1954) for studies of chemical development (see § 11.3). Finally, large single crystal specimens, as free from imperfections as possible, are required for most of the physical investigations described in the previous section.

The properties of the interface between the silver halide and the surrounding medium are of fundamental importance for photographic sensitivity. Although the chemical behaviour and the nature of the reactions occurring at the interfaces may be deduced from analytical work with hydrosols, microscopic investigations are not possible. The large optically clear thin single crystal sheets of silver halides with plane parallel surfaces, which lie near the opposite extreme of the range of photographic systems, are ideally suited for the purpose. The crystals have stable surfaces, and with them it is possible to study primitive sensitivity, chemical sensitization, dye sensitization, the surface and the internal latent images and physical and chemical development by microscopic methods. The role of dislocations and crystal imperfections in photographic sensitivity has been clearly demonstrated and this could never have been accomplished by direct observations on silver halide hydrosols and fine grained liquid emulsions.

From the thin single crystal sheets, the normal photographic emulsion may be approached through two intermediate stages : first, large microcrystals grown from gelatin-free solutions, and then large microcrystals grown under more normal emulsion conditions. In this way the relevance of the conclusions drawn from work on large crystals of fused silver halide to the discussion of the properties of microcrystals in emulsions can be demonstrated.

There is thus a continuous range from the ultra-microscopic particles of the hydrosols to the large single crystal sheets, and the various facets of the complex photographic process are often thrown into sharper relief when they are studied in the different systems. The remarkable feature of current photographic research lies, not in the differences in the characteristics of the systems which are being found, but in the surprisingly uniform pattern of behaviour which is revealed. The two approaches to which reference was made in the introduction to this section are essentially complementary, and both the chemist and the physicist have still much to contribute to the advancement of knowledge of photographic sensitivity.

§4. THE PRIMITIVE SENSITIVITY OF SILVER HALIDE CRYSTALS

4.1. *General Remarks*

By 1947 it was well known that latent image is formed in the interior of crystals of silver halides in emulsions as well as on the surfaces (Berg, Marriage and Stevens 1941 a, b, Debot 1942, Hautot 1945, see also Berg 1947). In discussions of the formation of this internal latent image in terms of the Gurney–Mott theory, reference was often made to shallow internal traps for electrons, associated in a general way with internal crystal imperfections, without the nature of the traps or the imperfections being in any way defined. The properties of the shallow traps were contrasted with those of the deep surface traps which were supposed to be provided by specks of silver sulphide. This position was summarized by Berg (1948).

4.2. *The Primitive Sensitivity of Photographic Emulsions*

An important advance was then made by Lowe, Jones and Roberts (1951), with their work on the primitive sensitivity of a fine grain positive emulsion. They described procedures for selectively oxidizing such chemical sensitizations as did not arise simply from contact between gelatin and silver halide and for normalizing the emulsion by removing excess oxidizing substances without sensitizing it. With these procedures they were able to take this particular emulsion through a series of sensitizing, oxidizing and normalizing cycles and to restore it to the same state of primitive sensitivity after each cycle. Their contribution lay in establishing that the primitive internal sensitivity of this emulsion was a characteristic and reproducible property ; the fact, which they also demonstrated, that the internal sensitivity of an emulsion in the primitive or chemically non-sensitized state is greater than the surface sensitivity was already known from other work. They attributed the primitive internal sensitivity, which resisted oxidation with aqueous solutions of halogens, to the presence of flaws, cracks and boundaries in the microcrystals. A similar view was expressed by Hautot and Sauvenier (1951 b).

4.3. *The Primitive Sensitivity of Silver Halide Crystals*

Hedges and Mitchell (1953 a, b) and Evans, Hedges and Mitchell (1955) then showed that the primitive sensitivity of large silver bromide crystals arose from the presence of dislocations and that the internal latent image was formed by the separation of silver atoms along the dislocation lines. Clark and Mitchell (1956), who extended this work to crystals of silver chloride and to mixed crystals of silver halides, reached the same conclusion. They found that the density of dislocations was greater in silver bromo-iodide mixed crystals than in pure silver bromide crystals. The possibility that other more macroscopic and less well defined internal imperfections, which also introduced internal surfaces, contributed to the internal sensitivity was never excluded, but no clear evidence for this was provided by the experimental work on lightly annealed single crystal sheets with well-defined sub-structures.

4.4. *Dislocations in Silver Halide Microcrystals*

Three methods were used by Evans and Mitchell (1955) for demonstrating the presence of dislocations in large tabular microcrystals of silver bromide (§ 3.2):

(1) Visible amounts of silver formed along lines on the surfaces during partial development after exposure, and located screw dislocations lying in planes passing through the centres of the microcrystals, parallel to the surfaces.

(2) When the crystals were treated with a number of silver bromide solvents, etch pits appeared where these screw dislocations met the edges. Shallow depressions, probably associated with edge dislocations, appeared in the centres of the tabular crystals under other conditions of partial dissolution.

(3) Exceedingly thin, straight filaments of silver were formed where the screw dislocations met the edge when the crystals were developed with a solvent-free physical developer after a short exposure. It seemed that surface latent image specks were formed at the ledges where the screw dislocations emerged, and that the deposition of silver upon them resulted in the growth of filaments with axial screw dislocations.

From these and other observations, Evans and Mitchell deduced that the crystals grew upon dendritic nuclei with six arms along $\langle 100 \rangle$ directions and postulated that the rotation of three arms with respect to the remaining three on a (111) plane about a [111] direction introduced a system of at least three coplanar screw dislocations. They proved experimentally that such rotations were possible and in this way were able to provide an explanation for the preponderance of tabular forms among microcrystals of pure silver bromide grown in the presence of excess bromide ions. It is possible that at least a fraction of these tabular crystals grow upon twinned nuclei produced by the rotation of the dendritic nucleus on a (111) plane through an angle of 60°. The re-entrant edges grow at a higher rate and a tabular crystal of triangular form results. The nucleation of new growth layers along the edges then occurs either at the re-entrant corners or at the points of emergence of screw dislocations. The occurrence of twinning on (111) planes in at least a fraction of the triangular and hexagonal crystals provides the only reasonable explanation of distributions of photolytic silver observed by Evans and Mitchell (1955) and by Fujisawa, Mizuki and Kubotera (1957).

These mechanisms are not applicable unless the crystal nuclei have the rock-salt structure. They cannot account for the growth of silver bromo-iodide crystals in tabular forms in emulsions if the nuclei, which are rich in silver iodide, have the hexagonal crystal structure which has been found by Chateau and Pouradier (1954), Raynaud and Pouradier (1955 a) and Berry and Marino (1955) in the initial precipitates produced when silver nitrate is run into mixed halide solutions in the absence of gelatin.

Two methods are employed for producing silver bromo-iodide emulsions for photographic purposes, which are distinguished by the way in which the silver nitrate and the alkali halide solutions are mixed together. In the first, which is used for preparing nuclear track emulsions (Demers 1950, 1952, 1954, Jenny 1951), the silver nitrate and the halide solutions are run simultaneously into the mixing vessel. The silver iodide is then distributed uniformly throughout the crystals, and cubic grains of uniform size and sensitivity are formed. In the second method, which is used for preparing the majority of emulsions, the silver nitrate solution is run at a controlled rate into the mixed halide solution. The nuclei which form are then rich in silver iodide compared with the silver halide which is subsequently deposited upon them. The crystals in these emulsions usually have triangular and hexagonal forms bounded by apparent (111) planes and sizes and sensitivities covering a wide range which determines the shape of the characteristic curve. All the evidence suggests that they are less perfect than the cubic crystals in emulsions prepared by the first method, and that their greater imperfection is responsible for their greater sensitivity.

The grains are, however, too small to allow any estimates of their state of imperfection to be derived from observations with the optical microscope. Herz and Grounsell (1957) have recently applied the micro-beam x-ray diffraction techniques developed by Hirsch and Kellar (1951) and Gay, Hirsch, Thorp and Kellar (1951) to demonstrate the presence of sub-structures in the microcrystals of many commercial silver bromo-iodide emulsions which they found to contain as many as twelve slightly misoriented polyhedral elements. This is valuable evidence to support the extension of the conclusions on the role of imperfections in photographic sensitivity, which have been based on observations with thin sheets and large microcrystals of silver halides, to the microcrystals of the technologically important silver bromo-iodide emulsions.

There are many ways in which dislocations and other internal imperfections may be introduced into silver bromo-iodide microcrystals:

(1) The iodide nuclei grow at a high initial rate under conditions of high supersaturation. They are usually imperfect, even when considered as crystals of silver iodide, and perfect crystals can never grow upon imperfect foreign nuclei.

(2) There is often a sharp concentration gradient in silver iodide between the nucleus and the outer parts of the crystal, and the associated internal strains may be relaxed by the introduction of a limited number of edge dislocation lines (Mitchell 1952, 1957 a, b). According to Chateau and Pouradier (1954) the iodide-rich nuclei may have the hexagonal structure and be transformed into silver bromo-iodide with the rocksalt structure at a later stage, with the introduction of internal strains.

(3) Even when perfectly formed cubic nuclei are produced and the iodide ions are uniformly distributed, as may be the case with the nuclear track emulsions considered earlier in this section, there will be a steadily increasing accumulation of lattice strains with increasing crystal size due to the incorporation of foreign ions and organic substances and to fluctuations in the uniform distribution of iodide ions. When these strains exceed a certain limit, corresponding with a

particular size of the microcrystals, they may be relaxed by the introduction of dislocation lines (Mitchell 1952). On this basis, the larger microcrystals will always tend to have a higher dislocation density, a higher primitive sensitivity and a higher capacity for chemical sensitization than the smaller microcrystals.

(4) From observations with the electron microscope, Arens (1948) concluded that the aggregation of three or six colloidal particles of silver halide could produce a composite nucleus upon which a tabular triangular or hexagonal crystal would grow. No evidence for the operation of this mechanism has been found in any work undertaken at Bristol. Hautot and Sauvenier (1955, 1957 b) have, however, expressed the opinion that the internal latent image is formed upon the contact surfaces between the aggregated elements of emulsion microcrystals and have made no reference to dislocations in their work.

4.5. *Sensitivity Centres*

The inadequacy of any discussion of photographic sensitivity based only upon the assumption of Frenkel and Schottky disorder in otherwise perfect silver halide microcrystals became evident during 1950 as a result of the work of Lowe, Jones and Roberts (1951) on primitive sensitivity which was presented at the Bristol Conference and of the work on the properties of thin sheet crystals of silver halides which was begun at Bristol at that time and showed almost immediately that internal sensitivity was determined by the presence of crystal imperfections. Two new hypotheses concerning primitive sensitivity were therefore introduced in a lecture given in April, 1951, in the University of Liège (Mitchell 1951 a), and in a paper read at the Paris Conference in September, 1951 (Mitchell 1952):

(1) The primitive internal sensitivity of microcrystals in emulsions is due to the presence of edge dislocations.

(2) The first stage in the formation of the latent image is the combination of an interstitial silver ion and an electron with a silver ion at an exchange site on the external surface or along an edge dislocation line. Surface exchange sites are now usually described as kink sites (Burton and Cabrera 1949 a, b) and exchange sites along edge dislocation lines as jogs (Seitz 1950 a). The result is the adsorption of a silver atom at a site adjacent to a silver *ion*† and this is now described as a latent pre-image speck (Mitchell 1957 a, b, see § 7.4.1).

At the same time, Mitchell pointed out that the areas around the points of emergence of dislocations through the surface of the crystal would have enhanced chemical reactivity on account of the high local surface density of kink sites. The chemical sensitizers would therefore be preferentially adsorbed at these sites and react to give sensitizing products. According to this concept, both the surface and the internal latent images are formed at dislocations, introduced during the nucleation and growth of the silver halide microcrystals. Such a topochemical relationship between the surface and internal latent images had not previously been considered and Mitchell introduced the term ‘ sensitivity centre ’ to describe the internal imperfection or dislocation together with its associated area of enhanced surface reactivity (Mitchell 1952).

The role of edge dislocations in the formation of the internal latent image was also discussed at this time by Seitz (1951, 1952, 1956). He based his approach on the Gurney–Mott theory and postulated that jogs on dislocation lines occupied by silver ions would furnish efficient traps for conduction electrons at room temperature. These silver ions have an effective half-unit positive charge. The trapping of an electron produces a silver atom, and a half-unit negative charge appears at the jog which then attracts an interstitial silver ion to the site to restore the effective half-unit positive charge and thus reset the trap. This mechanism encounters the same difficulty as any other Gurney–Mott mechanism. Positive holes have a higher mobility than the interstitial silver ions, and if the electron is trapped as postulated by Gurney and Mott and by Seitz, the localized negative charge which results is more likely to be neutralized by a positive hole than by an interstitial silver ion. If this occurs, the result is simply the recombination of an electron and a positive hole.

---

† See § 7.3.

Mitchell (1951 a, 1952) had considered this problem and avoided the difficulty. He concluded that the effective half-unit positive charge of a silver *ion* at a surface kink site or jog could not provide a sufficiently deep potential energy well to ensure the trapping of an electron at room temperature : the negative potential arising from the polarization of the surrounding medium reduces the depth of the trap. He therefore postulated that the electron is not trapped unless the depth of the positive potential energy well is increased by the presence of a mobile silver ion in the immediate neighbourhood. With this mechanism, a localized negative charge which can attract a positive hole never appears at the trapping site of the electron, and this feature distinguishes the description of the first stages in the formation of the latent image which has been given by Mitchell, and which will be summarized in §§ 7.4.1 and 7.4.2, from that given by Gurney and Mott and by Seitz. A mechanism with this property is essential to account for the quantum efficiencies of 0·4 to 0·5 which are found for the separation of silver atoms in plastically deformed pure silver halide crystals during exposure to light (see § 7.4.4 and § 10.4 for further discussion of this question).

Direct experimental evidence was then obtained by Mitchell and his co-workers which established the validity of the new concept of the sensitivity centre (see § 4.3 and § 4.4), and the properties of the sensitivity centres were then examined in a series of papers (Mitchell 1953, 1954, 1956, 1957 a, b), without any further modification of the essential features of the model.

Dislocations have thus three functions in photographic sensitivity :

(1) The growth of the crystals during Ostwald ripening may often be determined by the presence of screw dislocation lines (Evans and Mitchell 1955) and then proceeds according to the mechanisms proposed for the first time by Frank (1949 a, b, 1952 a, b). Silver atoms do not normally separate along screw dislocation lines to form internal latent image during exposure.

(2) The internal latent image is formed by the separation of silver atoms along edge dislocation lines introduced to relax internal strains in the crystals.

(3) The preferential dissolution of the crystal around the points in which dislocations of both types meet the surface produces areas of enhanced chemical reactivity due to the occurrence of a higher-than-average local surface density of kink and edge sites. This is important for chemical sensitization, which will be discussed in the next section.

§ 5. CHEMICAL SENSITIZATION

5.1. *Introduction*

There are two stages in the manufacture of a photographic emulsion which may, however, not always be clearly separated. In the first, to which brief reference was made in § 4.4, a suspension of silver halide microcrystals with a characteristic distribution of shapes, sizes and internal crystal imperfections is produced, usually in the presence of ammonium hydroxide and an excess of halide ions. If set, washed, re-melted and coated at this stage, the emulsion is, in general, found to have low surface and relatively high internal sensitivity. If the washed re-melted emulsion is then heated for a period at about 45°c, the surface sensitivity rises and the internal sensitivity, although at first unaffected, may later fall to a lower value. This process is called digestion or after-ripening. In this section, we shall be concerned with the changes which occur during digestion under different physico-chemical conditions.

The earliest method of chemical sensitization consisted simply in heating the emulsion, after the adjustment of the silver and hydrogen ion concentrations to suitable values, and relied upon the natural inherent sensitizing capacity of selected gelatins or upon unknown foreign sensitizing substances present in them. The great contribution made by Sheppard (1928) to photographic sensitivity was his recognition of the important role which compounds with labile sulphur atoms such as allylthiourea (1925 a, b) and sodium thiosulphate (1925 c, 1926) play during digestion. He concluded that sulphur sensitization led to the formation of specks of silver sulphide which were randomly distributed over the surfaces of the crystals. Soon afterwards, Steigmann (1928) expressed the opinion that the previously unrecognized sulphur sensitizer in an active gelatin was derived from cystine during the alkaline processing of the gelatin, and was not naturally occurring allylisothiocyanate, as had been suggested by Sheppard.

5.2. *The Sensitizers in Normally Processed Active Gelatins*

There has been a renewal of interest in the nature of the sensitizers in normally processed active gelatins due to the ease and efficiency with which mixtures of closely related substances may be separated and identified by the methods of paper partition chromatography and paper electrophoresis. In agreement with the conclusion reached by Steigmann (1951), Wood (1952, 1954) has established that the sulphur sensitizers in the gelatin extracts which he has investigated are inorganic thiosulphate and polythionate anions. He has found no organic sulphur compounds. Russell (1956) has confirmed the presence of thiosulphate and tetrathionate ions in aqueous gelatin extracts by polarographic methods. Krummenerl (1952, 1955) has also described the separation of sulphur sensitizers from normally processed gelatins by electrodialysis. The method of paper electrophoresis has been applied by Wood (1955 a) to the identification of the reducing substances in active gelatins. He has concluded that sulphite ions may be present.

5.3. *Methods of Chemical Sensitization*

5.3.1. *Adjustment of silver and hydroxyl ion concentrations and digestion.*

The simplest method of chemical sensitization consists in heating a washed emulsion made with a gelatin, freed as far as possible from compounds with labile sulphur atoms, after the adjustment of the silver and the hydroxyl ion concentrations to suitable values (Wood 1953, 1955). Loening (1951 a, b) and Bourne and Loening (1953) have also studied the spontaneous sensitization of silver bromide hydrosols with gelatin as a function of the silver and hydroxyl ion concentrations.

5.3.2. *Reduction sensitization.*

Although it had always been assumed that part of the surface sensitization arising from digestion was due to the formation of silver by the reduction of the silver halide, and reduction sensitization with sodium sulphite had previously been demonstrated by Carroll and Hubbard (1933), the first clear experimental distinction between reduction and sulphur sensitization was made by Lowe, Jones and Roberts (1951). They showed that the surface and internal sensitization resulting from the addition of stannous chloride to an inert gelatin before digestion was selectively destroyed by treatment with a dilute solution of potassium ferricyanide which did not affect sulphur sensitization. They thus placed the occurrence of reduction sensitization on a sound experimental basis. Evans, Hedges and Mitchell (1955) have used this method, and related methods with dilute solutions of chromic acid of carefully determined concentration as selective oxidizing agents, to establish the occurrence of reduction sensitization when large crystals and microcrystals of silver bromide are digested with gelatin solutions containing lactose, sodium sulphite, sodium formaldehyde sulphoxylate (see also Lowe and Jones 1948) and formaldehyde. The properties of emulsions, sensitized with reducing agents and with sulphur compounds, have also been studied by Hautot and Sauvenier (1953 b, c, 1957 a, b).

5.3.3. *Sulphur sensitization.*

Stevens (1953 b) has described the successful sensitization of photographic emulsions with dilute solutions of hydrogen sulphide which had previously been stated to cause only fog, and Loening (1953) has been able to sensitize silver bromide hydrosols in the same way. Evans, Hedges and Mitchell (1955) showed that large crystals of silver bromide could be effectively sensitized simply by exposing them to moist hydrogen sulphide carried at a low partial vapour pressure in a stream of nitrogen. These are the simplest methods of sulphur sensitization.

The methods which are of technological importance depend upon the adsorption of organic substances and inorganic anions with labile sulphur atoms and the breakdown of the adsorption complexes to silver sulphide (Carroll and Hubbard 1931, 1934). James and Vanselow (1951, 1954) have demonstrated the adsorption of allylthiourea molecules and of thiosulphate ions on the surfaces of silver bromide particles by the displacement of adsorbed dyes in chromatographic columns filled with silver bromide. They have introduced a valuable two-stage method for producing less than a monolayer of silver sulphide on the surfaces of silver bromide crystals in which a monolayer of allylthiourea is first adsorbed and the adsorption complex then converted into silver sulphide by raising the *pH* of the solution (James and Vanselow 1953 a, b). With this method, Mitchell (1957 a, b, d) has been able to demonstrate preferential sensitization with very dilute solutions of allylthiourea of the areas around the lines in which subboundaries and dislocations meet the surfaces of lightly annealed sheet crystals of silver bromide.

The properties of sulphur-sensitized photographic emulsions have been studied by Hautot and Sauvenier (1953 b, c, 1957 a) and by Eggert and Okamoto (1956 a, b).

5.3.4. *Gold sensitization.*

This important method of chemical sensitization was discovered in the Agfa Research Laboratories in Wolfen in 1936 by Koslowsky (Waller, Collins and Dodd 1943, Mueller 1949, 1955, Koslowsky 1951, 1953, Steigmann 1952 a, b, 1953 a, b, Amman-Brass 1953, Hautot and Sauvenier 1953 a, 1957 a). It is usually achieved by the addition of complex salts of monovalent gold to photographic emulsions before digestion with sulphur sensitizers, and results in a substantial increase in the sensitivity of the emulsion for incomplete development (as is usually the case in standard procedures). The sensitivity of x-ray emulsions is also increased by gold sensitization (Hoerlin and Mueller 1950). Evans, Hedges and Mitchell (1955) and Clark and Mitchell (1956) found that the surfaces of silver halide sheet crystals were effectively sensitized simply by immersion in an aqueous gelatin-free solution of sodium aurous di-thiosulphate which probably produces an adsorbed monolayer of aurous sulphide upon them. Low intensity reciprocity failure (§ 8) was reduced by the addition of potassium chloroiridite to the gold sensitizing solution.

The use of other metallic salts or compounds for increasing speed (Larsen, Mueller and Hoerlin 1953, Mueller 1955), modifying the characteristics of an emulsion from the point of view of reciprocity failure (Stauffer and Smith 1948), or stabilizing gold-sensitized emulsions (Yutzy and Leermakers 1949, Damschroder 1949) has been described in the patent literature.

Beyond this, further increases in the effective speed of high-speed negative emulsions are achieved by controlling the dynamic adsorption equilibria at the silver halide–gelatin interface during development, by the addition of surface active agents which act as development accelerators. Some reference will be made to this subject in §§ 11.2.4 and 11.4.1.

5.4. *The Nature of Chemical Sensitization*

When an emulsion is washed at the end of the crystal growth stage in its manufacture to remove ammonium hydroxide and excess alkali halide, localized dissolution, which results in the formation of terraced depressions, always occurs where dislocations and other internal imperfections meet the external surface. The edges of the incompleted planes of ions which form the terraces are irregular, due to the occurrence of a high linear density of kink sites along them (see Burton and Cabrera 1949 a, b). The surfaces of the sensitivity centres are thus characterized by a high local density of reactive kink and edge sites ; they provide the most reactive surface areas of the crystals, and the products of chemical sensitization are therefore formed at the sensitivity centres with higher probability than on the remaining surfaces (Mitchell 1954, 1956, 1957 a, b).

Some partial dissolution and rearrangement of the {111} surfaces of the microcrystals must also occur during the washing of the emulsion which precedes digestion. Although the {111} surfaces may be in equilibrium in the presence of ammonium hydroxide and an excess of halide ions, they should, according to the ideas of Kossel (1927) and of Stranski (1928), be replaced during washing by finely faceted surfaces with tiny cube corners bounded by {100} planes. This conclusion, which is based upon the analysis of the surface properties of an ideal ionic solid such as potassium chloride, is probably also applicable to silver halide crystals in aqueous solutions. Terraces formed by incompleted planes on the surfaces of the facets are important for the discussion which follows.

Digestion at relatively high silver and hydroxyl ion concentrations (§ 5.3.1) produces, at the sensitivity centres, a distribution of very small groups of silver atoms which are surrounded by gelatin molecules (Mitchell 1953 b, 1957 a, b, see also §§ 5.4 and 5.6). A similar result is achieved by reduction sensitization. The lack of adhesion of the silver to the surface has been demonstrated in the case of large silver bromide crystals by stripping the moist gelatin film from the surface and removing the bulk of the surface sensitization with it (Mitchell 1953 b, 1957 a, b). It is also probably demonstrated by the loss of surface speed observed by Wood (1955 b) to follow the adsorption of dyes at the surfaces of the microcrystals in his sensitized emulsions. Many dyes are known to displace gelatin from the silver halide surfaces (Pouradier and Roman 1952, West, Carroll and Whitcomb 1952, 1953), and its displacement should carry the sensitizing groups of silver atoms further away from the surface.

The surface speed achieved by sulphur sensitization, or by combined sulphur and reduction sensitization, or sulphur and gold sensitization, is not reduced by stripping the gelatin film from the surfaces of large crystals or by the dye sensitization of emulsions (Mitchell 1953 b, 1957 a, b, Wood 1955 b). It arises from the formation of silver sulphide which is strongly adsorbed at the surfaces of the sensitivity centres.

The heats of adsorption of the sensitizing ions and molecules are greatest when their labile sulphur atoms are adsorbed at sites adjacent to surface kink sites occupied by silver ions. The adsorption of a thiosulphate ion, for example, at such a site is accompanied by the passage of two halide ions into solution and the formation of an adsorbed silver thiosulphate molecule. This molecule then reacts with a hydroxyl ion to form a silver sulphide molecule and a singly charged sulphate ion. In this way, the kinked edges of the surface terraces become 'decorated' with adsorbed molecules of silver sulphide. The continuation of this process leads to the formation of monolayers of adsorbed molecules of silver sulphide over the surfaces of the terraced depressions at the sensitivity centres. The localization of the silver sulphide at the crystal imperfections thus results from the preferential adsorption of the sulphur sensitizer at surface kink and edge sites and its reaction at these sites to give adsorbed molecules of silver sulphide. This is the reason why the surface latent image (see § 4.5) is formed at the same sites as the internal latent image (Evans and Mitchell 1955, Mitchell 1956). No evidence has been found for the formation of randomly distributed sensitivity specks of silver sulphide on the surfaces of well-formed pure silver bromide crystals as would have been anticipated from the conclusion of Sheppard, Trivelli and Wightman (1927).

Under most practical conditions of sulphur sensitization, surface sensitivity rises and low intensity reciprocity failure falls when digestion is prolonged due to the formation of silver in increasing amounts in addition to silver sulphide. Both the optical absorption and the sensitivity of the emulsion are then extended to longer wavelengths (§ 5.6). That silver is formed is demonstrated by the disappearance of these additional properties when the emulsion is treated with a dilute solution of chromic acid for a time during which sulphur sensitization is unaffected.

In a series of model experiments, Evans, Hedges and Mitchell (1955) have shown that the surface sensitivity of silver bromide crystals, upon which a monolayer of silver sulphide has been formed, is increased by the deposition of about one-tenth of a monolayer of silver atoms. Crystals sensitized in the absence of gelatin in this way fog rapidly on storage. The stability of the sensitization is increased if the crystals are immersed in a solution of gelatin immediately after the deposition of the silver and dried. The silver atoms, which are then located at the interface between the silver sulphide and the gelatin, are far more stable than the adsorbed silver atoms of latent pre-image specks, formed by the action of light, which will be discussed in § 7.4.1. Unlike the latent pre-image specks, they cannot dissociate as a result of relatively low energy thermal activation into interstitial silver ions and conduction electrons. Some silver atoms, co-adsorbed with silver sulphide monolayers and stabilized by association with gelatin molecules, may be formed at the sensitivity centres during prolonged digestion. The bulk of the silver which is formed at this stage of chemical sensitization is, however, not adsorbed to the silver sulphide : it is present as small groups of silver atoms associated with gelatin molecules.

The Russian workers (Chibisov, Titov and Mikhailova 1951, 1952, Chibisov 1956) place considerable emphasis on the aspects of chemical sensitization considered in the previous paragraph. They maintain that silver sulphide as such is not important for chemical sensitization and that the presence of sulphur compounds simply facilitates the formation of the sensitizing product. This they assume to be silver present in the sub-surface layers of the crystals in the form of F-centres. The experimental results which have been obtained at Bristol do not support this view. They show that effective surface sensitization of silver bromide crystals can be achieved in conditions where no product but silver sulphide can be formed and where no diffusion of this silver sulphide into the sub-surface regions of the crystals can possibly occur. The Russian viewpoint corresponds with the earlier theory of reduction and sulphur sensitization which was worked out by Mitchell (1948, 1949 a, b, 1951 a, 1953 a, see also Pick 1951, Berg 1953, Staude 1956) on the basis of the experimental work of Stasiw and Teltow (§ 10.2.3 (*b*)). In his last contributions on this theory, Mitchell (1951 a, b, 1953 a) suggested that reduction sensitization resulted in the formation of aggregates of two adjacent F-centres, and sulphur sensitization in the formation of association complexes of divalent sulphide ions and vacant halide ion lattice sites in the sub-surface regions of the sensitivity centres. This theory has since been developed and extended by Stasiw (1950, 1951, 1952 a, b, 1953, 1954, 1955, see also Pick 1951 and § 10.2.3 (*b*)). Hautot and Sauvenier (1953 a, b, 1957 a, b) have expressed the opinion that a sub-surface complex, in which a divalent sulphide anion with an adjacent F-centre is associated with an interstitial silver ion, is produced by combined sulphur and reduction sensitization. Mitchell (1951 a, b) had postulated that this particular complex was formed in the first stage of the formation of the latent image in sulphur-sensitized crystals, and Stasiw (1953) associated it with adsorption bands which appear during the exposure of mixed crystals of silver halides and silver sulphide at low temperatures.

The author has been largely responsible for the introduction of descriptions of chemical sensitization depending upon the participation of vacant halide ion lattice sites. He must therefore state that he is now of the opinion that the experimental evidence which is available on the sulphur sensitization of silver halide crystals by the methods used in photographic technology is best interpreted in terms of the simple co-adsorption of silver sulphide or aurous sulphide molecules and silver or gold atoms at the surfaces of the sensitivity centres and does not require their subsequent diffusion into the sub-surface regions or their being built into the crystal by the deposition of layers of silver halide over them (Mitchell 1953 b, 1955 a). This does not imply that the theory, as developed by Stasiw (1955), is not applicable to the discussion of photochemical changes in mixed crystals of silver halides with silver sulphide in the *unstable* state which results from their being quenched from above 400°c to room temperature (see § 10.2.3 (*b*)).

When potassium aurous thiocyanate is added to the emulsion before digestion, at least a fraction of the adsorbed molecules of silver sulphide are converted into molecules of aurous sulphide and any co-adsorbed silver atoms are converted into far more stable gold atoms. Groups of silver atoms associated with gelatin molecules are partially converted into groups of gold atoms. This summarizes the present working hypotheses on the nature and distribution of the products of chemical sensitization. The functions of chemical sensitization will be discussed in § 7.

5.5. *Antifoggants and Stabilizers*

That active sites such as kink sites are involved in chemical sensitization is demonstrated by the effect of minute traces of many strongly adsorbed antifoggants as, for example, 2-mercaptobenzimidazol, in reducing the rate of sulphur sensitization or in completely preventing its occurrence under otherwise favourable conditions. It may be deduced that this class of antifoggants (Birr 1952 b, 1953, 1954 a), by their own adsorption, prevent the adsorption of the sulphur sensitizers at sites adjacent to surface kink sites occupied by silver ions.

An important function of antifoggants and stabilizers is that of reducing the rate at which fog develops during the storage of an emulsion, due to the continuation of the reactions which result in the formation of silver atoms at the adsorbed silver sulphide monolayers of the sensitivity centres. It will be seen in § 7 that a group of two silver atoms already constitutes a latent sub-image speck, so that an emulsion which has received combined sulphur and reduction sensitization or gold sensitization is particularly susceptible to the formation of fog on storage.

The problems of stabilization have been discussed by Birr (1952 b), who has been responsible for the development of a new class of particularly effective stabilizers—the azaindolizines—which are not strongly adsorbed to silver halides and which are superior for these applications to the heterocyclic nitrogen compounds which were previously employed (Birr 1952 a, 1954 b, 1955 a, b). Storage fog arises from the adsorption of silver ions by silver or gold atoms associated with the silver sulphide layer, either from interstitial positions in the crystal or from complexes formed between silver ions and gelatin molecules, followed by the transference of electrons from reducing molecules and the formation of latent sub-image specks. The most effective stabilizers are substances which prevent this from occurring, by forming more stable complexes with silver ions.

5.6. *Spectral Sensitization arising from Chemical Sensitization*

The fact that chemical sensitization extends the sensitivity of photographic emulsions to longer wavelengths has been known for many years. Capstaff and Bullock (1920) described how red sensitivity appears when emulsions are treated with sodium bisulphite, and Sheppard (1926) described further work by Bullock which established the appearance of a strong sensitivity band in the extreme red and infra-red for prolonged exposures following the treatment of photographic emulsions with very dilute solutions of sodium thiosulphate and the subsequent production of silver sulphide.

Work on this subject was begun in the Agfa Laboratories at Wolfen in 1938 (Eggert and Biltz 1938, 1939, 1940, Eggert and Kleinschrod 1939, 1940 a, b, Arens, Eggert and Kleinschrod 1943) and later continued by Biltz (1951, 1952) in the Kodak Laboratories at Rochester. It has established the close parallelism between the wavelength dependence of the optical absorption beyond the long-wave edge of the lattice absorption band and the photographic sensitivity of chemically sensitized emulsions in the same wavelength region. As already mentioned in § 2.11, Amsler (1953) found the same wavelength dependence for the photocurrents in his silver-sensitized silver bromide photocells. This suggests that part of the absorption beyond the lattice absorption band which increases in chemically sensitized emulsions during digestion is due to the production of silver. This silver must be in a form in which it cannot initiate physical or chemical development. It is probably present as very small uncharged groups of silver atoms mechanically associated with gelatin molecules within 40 Å of the surfaces of the sensitivity centres and not actually in contact with the silver halide.

Silver sulphide and co-adsorbed silver atoms at the surfaces of the sensitivity centres would also contribute to the optical absorption at longer wavelengths. Hautot and Sauvenier (1953 a, 1957 a, b) believe that the long wavelength absorption is due to complexes of the type mentioned in § 5.4, formed in the sub-surface regions of the crystals. Loening, Sutherns and Bourne (1954), on the other hand, attribute this absorption to silver hydroxide complexes and to complexes formed between silver ions and gelatin molecules ; such complexes are, however, unlikely to be present in any significant concentration in fully stabilized commercial photographic emulsions.

§ 6. SPECTRAL SENSITIZATION WITH DYES

6.1. *Experimental Observations on Dye Sensitization*

Chemically non-sensitized, pure silver bromide emulsions have an exceedingly small optical absorption and photographic sensitivity at wavelengths greater than 4900 Å (Hautot and Sauvenier 1957 a). Both optical absorption and photographic sensitivity are extended to 7000 Å by chemical sensitization during prolonged digestion, as has been described in § 5.6.

The most important method of spectral sensitization depends, however, upon the adsorption of monolayers of suitable dyes upon the surfaces of silver halide microcrystals. The dye molecules absorb light quanta and transfer the energy to the crystal. Photochemical changes result which would not be caused by quanta of the same energy in the absence of the dye, and the interpretation of this aspect of dye sensitization presents interesting theoretical problems. Spectral sensitization with dyes acts independently of the different forms of chemical sensitization which were discussed in §§ 5.3.3 and 5.3.4, and the distribution of the latent image in chemically sensitized crystals is not affected by dye sensitization. Many important advances have been made in this field during the past ten years, and reviews have been published by Carroll and West (1951), Scheibe (1952), West and Carroll (1954), Dickinson (1954) and Mees (1955).

The cyanines occupy a predominant position in the field of spectral sensitizers. These widely used dyes have a coplanar arrangement of basic heterocyclic nuclei linked by a methine or polymethine bridge. Much progress has been made in establishing the relationship between the structure of the dye molecules and their absorption spectra in dilute solutions. The empirical relationship between the optical absorption, the nature of the heterocyclic nuclei and the length of the bridge has been extensively investigated by Brooker and his co-workers and the essential factors which determine the positions of the optical absorption bands are now well understood (Brooker, White, Sprague, Dent and van Zandt 1947, Brooker 1954, 1955).

The importance of the coplanarity of the heterocyclic nuclei and the methine or polymethine bridge for efficient sensitization which was noted by Sheppard, Lambert and Walker (1941) has been demonstrated by the synthesis of many dyes in which the molecules are distorted by repulsive interactions between substituent groups (Brooker, White, Heseltine, Keyes, Dent and van Lare 1953). These dyes are found to be inefficient sensitizers or even non-sensitizers.

Dye molecules are not effective as spectral sensitizers unless they are adsorbed upon the surfaces of the silver halide microcrystals. West, Carroll and Whitcomb (1952, 1953) have measured the heats of adsorption of a number of sensitizing dyes and demonstrated that many dyes displace adsorbed gelatin.

Although silver halide microcrystals are efficiently sensitized by the adsorption of isolated dye molecules, the practically important spectral sensitization is due to the formation of aggregated monolayers in which the long, planar dye molecules are adsorbed in the same way as a spaced pack of cards resting on their long edges and tilted at an angle between 59° and 70° to the surface (Sheppard, Lambert and Walker 1939, West and Carroll 1951, Carroll and West 1951). These aggregated monolayers which cover considerable areas with dye molecules in parallel orientation play an important role in spectral sensitization.

The absorption spectra of the dye molecules are changed as a result of their adsorption. Instead of the broad absorption bands which are characteristic of dilute solutions or of small surface coverages, strong, very sharply defined absorption bands appear (Scheibe 1948, Dickinson 1948, 1950, Carroll and West 1951). This unusually selective absorption is of great importance for many applications, as, for example, in the sensitization of emulsions for use in integral tri-pack methods of colour photography.

There is, however, one serious disadvantage associated with the adsorption of the dye in the aggregated state : the sensitizing efficiency is greatly reduced. The dyes are then often strongly fluorescent and the absorbed energy is re-emitted instead of being transferred to the crystal to cause photochemical changes. This phenomenon of resonance fluorescence, which was first observed by Jelley (1936, 1937) and by Scheibe (1936, 1937 a, b, 1938), is almost unique in liquid and solid systems.

In these circumstances, the sensitizing efficiency may be restored to a value nearer that found for isolated dye molecules by the co-adsorption either of other dyes or of related molecules which are called supersensitizers (West and Carroll 1951, 1953). The supersensitizers quench the fluorescence and increase the utilization of the absorbed energy for the production of photochemical changes. The combination of dye and supersensitizer is essential whenever a dye is employed in the aggregated state, and many combinations are cited in the patent literature. West and Carroll (1951, 1953) have found that non-sensitizing, non-planar cyanine dyes, co-adsorbed with sensitizing dye molecules in aggregated monolayers, may quench the fluorescence and reduce the sensitizing efficiency to a negligible value. These dyes are described as antisensitizers. Whereas supersensitizers increase the efficiency of sensitization by promoting the transference of the excitation energy from the dye molecules to the crystal, antisensitizers appear to absorb and to degrade the excitation energy into thermal energy. The antisensitizers do not affect the sensitivity of the emulsion to radiation with wavelengths in the lattice absorption band. In this respect they differ from the desensitizers, and from many dyes which show strong desensitizing properties, particularly when adsorbed in the aggregated state, and depress the surface sensitivity over a wide range of wavelengths (Spence and Carroll 1948, Dickinson 1954).

The close parallelism between the optical absorption of the adsorbed dye and the spectral sensitivity of the emulsion was established by Leermakers, Carroll and Staud (1937). Since then, the experimental work on adsorption, the spectral absorption of the adsorbed dye (West and Saunders 1957) and spectral sensitization has been supplemented by measurements of the photoconductivity of dye-sensitized emulsions (West and Carroll 1947, West 1951, West and Carroll 1951). West and Carroll have demonstrated the occurrence of photoconductivity when dye-sensitized emulsions are irradiated in the absorption band of the dye, and have thus established that electrons are liberated in the crystal when light quanta are absorbed by the dye. The photocurrents are increased by the co-adsorption of a supersensitizer and decreased by an antisensitizer. The very close parallelism between the magnitude of the observed photocurrents, in the presence of both supersensitizers and antisensitizers, and the photographic sensitivity leaves little doubt that the release of conduction electrons is an essential feature of dye sensitization. The only puzzling result of this experimental work is the observation of West and Carroll (1947) that the photocurrents are not decreased by treating the emulsions with desensitizing dyes which reduce the photographic surface sensitivity to a negligible value. Sheppard, Vanselow and Happ (1940) and Amsler (1953) have also shown by their studies of the photo-voltaic effect that electrons are liberated when dyed silver–silver bromide electrodes are illuminated in the absorption band of the dye.

The photocurrents in dye-sensitized and non-dye-sensitized emulsions have the same wavelength dependence as the photographic sensitivity of these emulsions, and dye-sensitization does not alter the relative amounts of surface and internal latent image which are formed during exposure. An internal latent image is produced with high efficiency in large chemically non-sensitized sheet crystals upon the surfaces of which an aggregated dye and supersensitizer layer has been adsorbed (Clark and Mitchell 1956).

The cyanine dyes are not destroyed during exposure and this seems to require that dye sensitization should depend on the transference of excitation energy and not of electrons from the dye molecules to the crystal (Scheibe 1952, Mitchell 1957 a, b). If any other action of the dye molecules were involved, it would be anticipated that the unstable infra-red sensitizers would be most readily destroyed. Yet, even here, Eggert, Meidinger and Arens (1948) have shown that 160 silver atoms are produced per molecule of dye destroyed at 9500 Å and 8 per dye molecule at 10 500 Å. We therefore conclude that the experimental evidence which is available at the present time supports the hypothesis that dye sensitization results from the transference of excitation energy to the crystal rather than from the transference of an electron from the dye molecule as was suggested by Gurney and Mott (1938) but later withdrawn by Mott (1948). Nelson (1956 a, b, c) has recently expressed the opinion that his experimental work on the photoconductivity of adsorbed layers of cyanine dyes, and on the work functions of dye layers, supports the original Gurney–Mott viewpoint and has suggested that it was perhaps too readily abandoned. The same conclusion has been reached on similar grounds by Noddack, Eckert and Meier (1952), Noddack and Meier (1953) and Meier (1954). We are of the opinion that the weight of evidence favours the view expressed by Mott in 1948, but it is clear that the two alternatives still merit careful consideration.

6.2. *The Theory of Dye Sensitization*

A number of papers have been published on the interpretation of the absorption bands of the cyanine dyes. Kuhn (1948 a, b) has used the one-dimensional free-electron-gas model for the discussion of the dependence of the position of the absorption band on the length of the polymethine bridge, and Moffitt (1950) has given a quantum-mechanical discussion of the absorption process. The dye molecules have Σ_g ground states and the transition to the Σ_u excited state has a very high oscillator strength. Simpson (1947) suggested that a metastable triplet state was involved in dye sensitization and this possibility was also discussed by Umano (1956). No experimental evidence has yet been presented which requires the adoption of this viewpoint for its interpretation.

It now appears that the theory of dye sensitization proposed by Franck and Teller (1938, see also Franck and Livingston 1949) is correct in its essential principles. Franck and Teller applied the ideas of Frenkel (1931 a, b) on mobile excitation states, or excitons, in solids to the aggregated states of the pseudocyanines which had just previously been discovered by Jelley and by Scheibe. Scheibe (1938, 1948) had suggested that the aggregate behaved as an assembly of molecules through which the excitation state could be propagated until it either decayed with the production of resonance fluorescence or caused a photochemical change at a lattice imperfection. Franck and Teller proposed that the surface transition involved in the photochemical change was one which could be produced by the absorption of light with a wavelength beyond the long-wave limit of the lattice absorption band, but that the corresponding optical absorption was very weak. They suggested that the strong absorption band of the dye molecules in the aggregated state on the surface was associated with the creation of a freely mobile exciton and that the excitation energy could be transferred to the crystal at the imperfections to cause photochemical changes. This approach has also been followed by Mott (1948). Mitchell (1957 a, b) has recently given a more detailed discussion of dye sensitization based on the same principles which will now be summarized.

The low energy transitions which are involved in dye sensitization occur between occupied electronic energy states with localized wave functions, associated with surface halide *ions* or products of chemical sensitization, and unoccupied energy states, also with localized wave functions, associated with surface silver *ions*. In the absence of adsorbed sensitizing dyes, the transition probability is vanishingly small because of negligible overlap of the electronic wave functions of the initial and final states due to the location of the *ions* and molecules which are involved at separated sites along the edges of the surface terraces. The dye molecules are adsorbed along and parallel to these surface terraces (Mitchell 1957 b), so that there is extensive overlapping between the electronic wave functions in their ground and excited states and those of the initial and final states of the surface transition. The result is that energy absorbed in the dye molecules can be transferred to the crystal with high probability to cause electronic transitions involving *ions* along the edges of the surface terraces.

The discussion so far given applies to isolated dye molecules adsorbed along the edges of surface terraces. We have next to consider sensitization by adsorbed aggregated monolayers of dye molecules. West and Saunders (1957) have studied the absorption spectra of aggregated dye monolayers on the surfaces of large silver bromide sheet crystals in polarized light, and observed large dichroic ratios. This would be consistent with the adsorption of the dye as long parallel filaments with the individual dye molecules at right angles to the axes of the filaments (Dickinson 1950). The nuclei for these filaments would be formed by the adsorption of dye molecules along and parallel to the edges of surface terraces so that the axes of the constituent filaments of the aggregated monolayers would be at right angles to these terraces. Scheibe and Kandler (1938, see also Dickinson 1950) have shown that the intense narrow absorption band, which is characteristic of the aggregated state, occurs when the electric vector of the incident radiation lies along the axes of the filaments. In this case, it is at right angles both to the individual dye molecules in the filaments and to the edge of the surface terrace at which the filaments are adsorbed. This relationship is not favourable for the production of photochemical transitions which involve an oscillating electric dipole moment parallel to the edge. The inefficient sensitization produced by the aggregated dye compared with isolated molecules of the same dye may thus be explained.

Many dye sensitizers displace gelatin from the surfaces of the silver halide crystals and are thus adsorbed in actual contact with the surface. The axes of the individual filaments in the aggregated dye layers are at right angles to the edges of the surface terraces which thus represent discontinuities where the dye monolayer on one side of the edge is displaced along the normal to the surface with respect to that on the other side by the thickness of one surface plane of *ions*. This displacement introduces a perturbation which increases the probability for the production of photochemical changes along the edge by the transference of the excitation energy from the dye. The efficiency is nevertheless low for the reasons which have already been given ; the co-adsorption of a supersensitizer is essential for efficient sensitization by an aggregated dye.

Both the supersensitizers and the antisensitizers are adsorbed along and parallel to the edge which produces the discontinuity in the aggregated monolayer of dye. In this position, the individual molecules of the supersensitizers localize the excitation energy precisely where it is needed to cause low energy photochemical changes and increase the efficiency with which it is transferred to the crystal. The antisensitizers absorb and degrade the excitation energy in these same positions. The low optimum molecular ratio of supersensitizer to dye sensitizer is readily understood on this basis. The supersensitizer has only to be adsorbed where it is required to perform its essential function, namely, at the linear discontinuities in the aggregated dye layer along the edges of surface terraces whereas the dye molecules cover the entire surfaces of the terraces. The changes which follow the transference of the excitation energy from the dye layer to the crystal will be considered in § 7.2.5.

§ 7. THE FORMATION AND PROPERTIES OF THE LATENT IMAGE

7.1. *General Observations*

7.1.1. *Theories of the formation of the latent image.*

The theories of the formation of the latent image which have been discussed during the past ten years fall into three groups, in which it is assumed that one of the following plays the most important role in the chemical and photochemical changes determining the sensitivity of a photographic emulsion :

(1) light-sensitive complexes between gelatin and silver ions,
(2) vacant halide ion lattice sites and F-centres,
(3) crystal imperfections and Frenkel defects.

Loening (1951 a, b, 1952) and Bourne and Loening (1953) have concluded, from the results of their experimental work with silver bromide hydrosols, that photochemical changes in light-sensitive complexes between gelatin and silver ions are involved in the formation of the latent image. The extent of such complex formation is a function of the hydroxyl ion concentration and they assume that this is partly responsible for the influence of this variable in photographic sensitivity. Loening, Sutherns and Bourne (1954) have extended the theory and proposed a dual mechanism in which photochemical changes in both surface complexes and the silver halide crystal are assumed to occur during exposure. Sutherns and Loening (1956 a, b) have shown that silver sulphide disappears through chemical reactions during the exposure of chemically sensitized liquid silver bromide emulsions, but that these reactions are prevented by the addition of halogen acceptors which do not, by themselves, produce surface sensitization. From this observation, they have concluded that the most important function of silver sulphide in photographic sensitivity is not that of reacting with liberated halogen.

The trapping of positive holes is, however, one of the principal functions of the silver sulphide according to the second group of theories (Stasiw and Teltow 1941, Mitchell 1948, 1949 a, b, 1951 a, b, 1953 a, Stasiw 1949 a, b, 1950, 1951, 1952 a, b, 1953, 1954, 1955, Hautot 1951 a, b, Hautot and Sauvenier 1953 a, 1957 a, b, Chibisoff 1956) in which it is postulated that the products of chemical sensitization consist of complexes, formed in the sub-surface regions of the silver halide crystals, between vacant halide ion lattice sites, F-centres, divalent sulphide ions and interstitial silver ions. Reference is made to these theories in § 5.4 and in § 10.2.3 (*b*). The author and his co-workers have found no experimental evidence which requires, for its interpretation, the introduction of the hypothesis that the products of chemical sensitization are formed in the sub-surface regions of the crystals. They have developed a third theory of the formation of the latent image, in which the trapping of positive holes by adsorbed monolayers of silver sulphide and the chemical and photochemical properties of crystal imperfections are emphasized (Mitchell 1951 a, 1952, 1953 b, 1954, 1956, 1957 a, b, d). The theory, which has a sound experimental basis, is summarized in this section.

Other contributions to the theory of the formation of the latent image have been made during this period by Mott (1948), Hautot (1950), Hautot and Sauvenier (1951 b), Seitz (1951, 1952, 1956, see § 4.5) and Baker (1954). Steigmann (1953 c, 1954, 1955, 1956) has called attention to his earlier contributions to the subject. Reviews have been published by Pick (1951), Vassy (1953), Arens (1954 a, b), Klein (1954), Luckey (1954), Stasiw (1955), Frieser and Klein (1955), Mitchell (1955) and Chibisoff (1956).

7.1.2. *Intermediate steps in the formation of latent image specks.*

The formation of latent image specks proceeds through a number of intermediate steps. Katz (1949, 1950) and Webb (1950 a, b) have shown that the smallest stable latent image speck consists of a pair of silver atoms. It is called a latent sub-image speck (Webb and Evans 1938 a, b, 1940, Berg 1948, 1952 a, b, c), and is converted into a latent image speck by the addition of at least one further silver atom. A latent image speck is therefore a group of three or more silver atoms. A latent sub-image speck is itself formed through an intermediate group, for which we have found it convenient to introduce the term ' latent pre-image speck ' (Mitchell 1957 a, b).

The surface and internal latent images consist of latent image specks which are distributed over the external and internal surfaces of the sensitivity centres. The internal latent image does not catalyse the reduction of the crystal in the developer until the overlying silver halide has been removed by the action of a solvent.

Latent pre-image specks are unable to initiate development. They have only a transitory existence during exposure. The latent sub-image is distinguished from the latent image by the longer induction period required for its development (for further details see §§ 8 and 11).

These properties were not known at the time when Gurney and Mott (1938) advanced their theory in which the discussion of the trapping of the electron by the sensitivity speck depended upon the introduction of physical properties of massive silver and silver sulphide. The facts that adsorbed molecules instead of specks of silver sulphide, and groups of less than ten instead of a hundred or more silver atoms, are involved in the formation of surface latent image specks have necessitated extensive revision of the Gurney–Mott theory.

7.1.3. *Observations on the theory of the formation of latent image specks.*

A number of difficulties have not been recognized and discussed in much published work on the theory of the formation of latent image specks. The important photochemical changes involved in the formation of the latent image occur at sensitivity centres with higher probability than elsewhere. As positive holes and electrons usually recombine at imperfections in other inorganic crystals with the degradation of the excitation energy to thermal energy, it is important to inquire why this does not happen in the silver halides. It is also important to explain why a conduction electron, released at the surface when energy is absorbed by a sensitizing dye molecule, combines with a silver ion at a distant site instead of simply recombining with the localized positive surface charge at its site of origin.

According to the Gurney–Mott principle (see Mitchell 1957 a, b), the motion of electrons and of silver ions is involved in the formation of the latent image. The next question which therefore arises is why these electrons and silver ions condense to form discrete groups instead of a distribution of isolated silver atoms. In their theory, Gurney and Mott (1938) assumed that discrete groups were formed because the sensitivity specks became negatively charged by trapping conduction electrons, and then attracted interstitial silver ions. We shall see that neither the adsorbed molecules of silver sulphide which are responsible for sulphur sensitization, nor the latent sub-image specks, can trap conduction electrons and that a different mechanism has to be introduced.

We have finally to explain why groups of silver atoms when once formed do not simply act as recombination centres for electrons and positive holes. The importance of this question can be appreciated by reference to any one of the energy level diagrams showing the relative positions of the occupied energy levels in the silver halide and in the silver speck or sensitivity speck which were used in the discussion of the formation of the surface latent image or of photolytic silver on the basis of the Gurney–Mott theory (see, for example, Berg 1948, p. 262). The positive hole represents an unoccupied level in the full band of the silver halide and, if the energy levels are in the positions shown in these diagrams, it is impossible to avoid the conclusion that it should be trapped by a higher, occupied energy level in the sensitivity speck or silver speck. The positively charged speck would then attract the electron so that recombination would result. The building up of the latent image into visible groups of silver atoms requires the introduction of mechanisms in which such recombination is avoided.

The problems and difficulties raised here will be discussed in this section and again in § 10.

7.2. *Energy Levels in Crystals of Silver Halides*

7.2.1. *Energy levels within the crystals.* (See figure overleaf.)

Many photochemical properties of silver halide crystals are most conveniently discussed in terms of the energy level diagram (Mitchell 1957 a, b). The valency electrons of the halide ions occupy a band of levels between C and D, with a width of about 3·5 ev, which is called the full band of the crystal. The depth AC of the highest occupied levels in this band is about 6 ev. AC can be determined experimentally from the wavelength dependence of the external photoelectric effect in crystals of pure silver bromide (Fleischmann 1930, Taft and Apker 1957). The occupied levels of the full band are separated from the lowest unoccupied levels by the forbidden band BC. BC is determined experimentally from the long-wavelength limit of the internal photoelectric effect (see §§ 2.3 and 2.5); it is found to be about 2·5 ev. The unoccupied band of energy levels AB is called the conduction band and, during their lifetimes, electrons with corresponding energies can diffuse freely in the crystal and drift in electric fields. AB is the electron affinity of the crystal which is 3·5 ev.

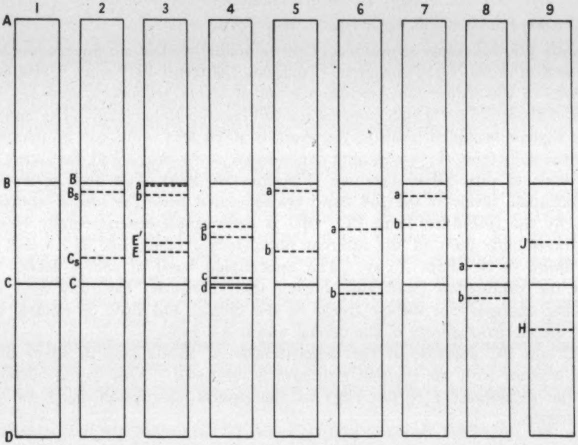
AC and BC decrease as the concentration of iodide ions in silver bromoiodide crystals with the rocksalt structure is increased. This would correspond with the displacement of the edge of the optical absorption band to longer wavelengths.

7.2.2. *The absorption of energy within the crystals.*

Energy may be absorbed within silver halide crystals in the form of either electromagnetic radiation or the kinetic energy of electrons or nuclear particles. The absorption of light quanta causes the transference of electrons from occupied levels in the full band CD to unoccupied levels in the conduction band AB, so that each absorbed quantum produces a positive hole and a conduction electron which can move independently in the crystal. The absorption of an X-ray or γ-ray quantum liberates a fast secondary electron which then transfers its kinetic energy to the crystal with the production of many pairs of positive holes and conduction electrons. The same process is involved when high-energy nuclear particles pass through the crystals.

In all these cases, the holes and electrons can diffuse through the crystals, and drift in electric fields to give rise to the observed photoconductivity and current pulses (§§ 2.5 and 2.6). In crystals of silver bromide, conduction electrons have a diffusion constant about 1 $\mathrm{cm}^2\,\mathrm{sec}^{-1}$ and a lifetime of about 10^{-5} sec at room temperature. The positive holes behave as polarons (Mitchell 1957 a, b, see also Allcock 1956). They have a much smaller diffusion constant—about 10^{-5} $\mathrm{cm}^2\,\mathrm{sec}^{-1}$ at room temperature.

The occurrence of optical absorption peaks within the main lattice absorption band of crystals of silver chloride and silver bromide has recently been established by Okamoto (1956) and by Tutihasi (1957). The absorption measurements have been made with thin films of the silver halides, deposited on quartz substrates, and the peaks, which are sharply defined at low temperatures, are attributed to the formation of excitons. Mitchell (1955, 1957 a, b) is of the opinion that excitons, created by the absorption of energy within the silver halide crystals,



Energy levels involved in the formation of the latent image in crystals of silver bromide

1. Energy levels within the crystals: AB, the conduction band; BC, the forbidden band; CD, the full band.

2. Localized energy levels at the surfaces of the crystals: BB_s, range of unoccupied levels associated with surface silver *ions*; C_sC, range of occupied levels associated with surface halide *ions*.

3. Localized levels associated with molecules of silver sulphide *adsorbed at sites adjacent to surface silver ions*: ab, acceptor levels which provide very shallow traps for conduction electrons; E′E, range of highest occupied levels which furnish traps for positive holes.

4. Localized levels associated with positively charged molecules of silver sulphide Ag_2S^+ adsorbed at sites adjacent to surface silver *ions*: ab, lowest unoccupied acceptor levels which provide traps for conduction electrons; cd, highest occupied levels. These molecules have a short lifetime and dissociate with the loss of a silver ion. It should be noted that the levels plotted in 3 and 4 are those for silver sulphide molecules adsorbed at sites adjacent to surface silver *ions* and therefore represent the deepest levels for these molecules.

5. Localized levels associated with latent sub-image specks Ag_2 adsorbed at sites adjacent to surface silver *ions*: a, acceptor levels which provide very shallow traps for conduction electrons; b, highest occupied levels which provide traps for positive holes.

6. Localized levels associated with positively charged latent pre-image specks Ag_2^+ adsorbed at sites adjacent to surface silver *ions*: a, acceptor levels which furnish traps for conduction electrons; b, highest occupied levels which are unable to trap positive holes. These specks have an extremely short lifetime and dissociate with the loss of a silver ion.

7. Localized levels associated with small uncharged latent image specks adsorbed at sites adjacent to surface silver *ions*: a, acceptor levels which furnish shallow traps for conduction electrons; b, highest occupied levels which provide traps for positive holes.

8. Localized levels associated with small positively charged latent image specks Ag_n^+ adsorbed at sites adjacent to surface silver *ions*: a, acceptor levels which provide effective traps for conduction electrons; b, highest occupied levels which do not provide traps for positive holes.

9. Energy levels of an adsorbed dye molecule: J, lowest unoccupied level; H, highest occupied level.

Note: No quantitative significance should be attributed to the positions of the localized energy levels in these diagrams. This purely qualitative representation is simply intended to illustrate the conclusions which have been reached regarding the mechanisms involved in the formation of the latent image. All the groups of localized levels extend over a wide range of energies but no quantitative experimental information is yet available.

do not cause any photochemical changes at room temperature which can be distinguished experimentally from those produced by the conduction electrons and positive holes into which they probably dissociate. No further reference will therefore be made to excitons in the discussion of the formation of the latent image.

The positive holes and the conduction electrons diffuse to the external surface and to crystal imperfections where they are trapped by localized occupied and unoccupied energy levels in the forbidden band. We have next to consider the nature of these surface levels.

7.2.3. *Surface energy levels.*

The energy required to remove an electron from a surface halide *ion* and to eject it from the crystal is less than AC. The surface halide *ions* introduce a group of localized levels above C in the forbidden band which are represented by CC_s in the diagram. The levels near C_s are associated with surface iodide *ions*. The energy released when an electron combines with a surface silver *ion* is greater than AB and the surface silver *ions* introduce a group of occupied levels BB_s below B in the forbidden band. The depth below B of the unoccupied levels in the group BB_s is not sufficiently great to ensure the trapping of conduction electrons by surface silver *ions* at room temperature (see § 7.3). The electrons have only a short lifetime in the traps before they are released again by thermal agitation and diffuse away (Mitchell 1952, 1957 a, b).

Occupied levels in the group CC_s provide deeper traps for positive holes than do levels in the group BB_s for electrons, and the lifetime of a positive hole appears to be determined by the time required for its diffusion from the point of creation to the surface of the crystal or to an internal dislocation. We assume that the positive holes are trapped before the electrons during the early stages in the formation of the latent image. Positive holes must be trapped at the external surface in order that halogen molecules can leave the crystal. The trapping of a positive hole results from the transference of an electron from a localized occupied surface level in the group CC_s to the unoccupied level in the full band. As a result of this, an (Ag^+X^-) group at a kink site becomes an (Ag^+X) group. A low-lying unoccupied localized surface level appears in the group CC_s which is associated with the excess silver ion at the kink site. The crystal is in an excited state determined by the presence of an electron in the conduction band and a low-lying unoccupied surface level in the forbidden band. We shall continue the discussion in § 7.3.

7.2.4. *Surface energy levels due to adsorbed chemical sensitizers.*

Molecules of silver sulphide adsorbed along the edges of surface terraces or as surface monolayers introduce localized occupied electronic energy levels in the forbidden band. These are represented by the group E′E in the energy level diagram. The width of this group of occupied energy levels, which is partly responsible for the extension of the optical absorption of chemically sensitized crystals to longer wavelengths, corresponds to the different sites and configurations of the adsorbed silver sulphide molecules. Neither the upper nor the lower limit has yet been determined, but the width of the group is not important for any of the arguments which will be presented.

Silver sulphide molecules have positive electron affinities and the higher unoccupied levels corresponding with the negatively charged states (Ag_2S^-) may also be plotted in the energy level diagram. Even when the polarization of the surroundings by the negatively charged molecule is taken into account, its electron affinity is unlikely to exceed 1·5 ev. We therefore reach the important conclusion that the unoccupied electronic energy levels lie above B, the bottom of the conduction band, and that adsorbed molecules of silver sulphide can therefore not trap conduction electrons (see also § 7.4.1 and captions 3 and 4 in the diagram).

The highest surface density of localized occupied energy levels in the forbidden band is furnished by the group E′E due to adsorbed molecules of silver sulphide which provide deeper traps for positive holes than the levels corresponding to halide *ions* on kink and edge sites. The trapping of a positive hole by a surface silver sulphide molecule leaves the crystal in an excited state with an electron in the conduction band and a low-lying unoccupied level in the group E′E corresponding to a positively charged Ag_2S^+ molecule. We shall again continue the discussion in § 7.3.

7.2.5. *Optical transitions involving levels in the forbidden band.*

We have seen in § 6.2 that optical transitions between localized occupied levels in the group CC_s and localized unoccupied levels in the group BB_s have a vanishingly small transition probability because of negligible overlapping of the electronic wave functions in the initial and final states on account of the separation of the kink sites, which are involved, along the edges of the surface terraces. The optical absorption is extended to longer wavelengths when molecules of silver sulphide are adsorbed at sites adjacent to kink sites occupied by silver ions, because the overlapping is greater and the transition probability correspondingly higher. The absorption of a light quantum in this wavelength region causes the transference of an electron from a level in the group E′E to BB_s from which it can be further excited to the conduction band by thermal agitation. This leaves an unoccupied level in the group E′E which corresponds with a positively charged Ag_2S^+ molecule.

The highest occupied electronic energy levels of the cations of the cyanin sensitizing dyes lie around H, between 7 and 8 ev below A (Mott 1948). The separation of H from the lowest excited level, J, is usually between 1·3 and 2 ev. The position of H shows that the dyes cannot trap positive holes. Neither can organic halogen acceptors which have occupied energy levels in a similar position. This explains why adsorbed dyes and halogen acceptors are not effective in promoting the formation of the surface latent image in chemically non-sensitized emulsions.

The excitation of a dye molecule by the absorption of a light quantum raises an electron from the level H to the level J. The vacated level H lies below localized occupied levels in the forbidden band which are associated with surface halide ions or adsorbed molecules of silver sulphide spatially overlapped by the dye molecules adsorbed along the edge of the surface terrace. An electron can therefore be transferred to the level H leaving an unoccupied level in one of the groups CC_s or E′E and conferring a negative charge on the dye molecule which raises all the electronic energy levels associated with it, including J. The electron is then transferred from J to a localized level in the group BB_s, associated with a surface silver *ion* from which it is raised to the conduction band AB by thermal agitation. The transference of the excitation energy from the dye molecule to the crystal thus raises an electron from an occupied to an unoccupied level in the forbidden band. This results in the appearance of a low-lying unoccupied level in the forbidden band associated with an Ag^+X or Ag_2S^+ surface unit and a conduction electron as in all the cases previously discussed.

The same excited state of a silver halide crystal thus results whether the energy is absorbed (i) within the crystal, (ii) by sensitizing molecules adsorbed at the crystal surfaces, or (iii) by sensitizing dye molecules adsorbed upon the crystal surfaces, followed by the transference of the excitation energy to the crystal.

7.2.6. *The sensitivity centres.*

The highest surface density of localized levels in the forbidden band associated with silver and halide *ions* at kink and edge sites and with adsorbed silver sulphide molecules occurs at the sensitivity centres, where the dye molecules are also adsorbed with highest probability. Although all the properties which are essential for photographic sensitivity are shown by the kinked edge of any surface terrace, it is clear that the photochemical changes which have been discussed in this section occur with higher probability at the sensitivity centres where there is a greater density of kink and edge sites than elsewhere on the surfaces. These changes leave the crystal in an excited state with an electron in the conduction band and a low-lying unoccupied level in the forbidden band. This level is associated with an excess silver ion adjacent to a halogen atom Ag^+X or with a positively charged silver sulphide molecule Ag_2S^+. The surface of the sensitivity centre thus becomes positively charged and a compensating negative charge is provided by conduction electrons. The question which immediately arises is why the conduction electrons are not trapped by the low-lying unoccupied levels with the emission of fluorescent radiation or the degradation of the excitation energy. If this were to occur to any extent, the photochemical process would have an extremely small quantum efficiency which is not the case.

7.3. *The Primary Stage in the Formation of the Latent Image*

We have seen that effective traps for electrons are not immediately available in crystals of silver halides in emulsions but that relatively deep traps are always available for positive holes. We have therefore discussed the trapping of the positive holes before the trapping of the electrons and we have now to explain why the trapping of the positive hole is not immediately followed by the trapping of the electron at the same site.

This possibility can be excluded only if the grouping which results from the trapping of the positive hole dissociates so that the low-lying unoccupied level is raised to a position below the bottom of the conduction band. The most probable dissociation mechanism is that the excess silver ion, the positive charge of which is equivalent to the charge of the positive hole, jumps into an interstitial position and diffuses away within the lifetime of the electron in the conduction band. The dielectric conditions at the interface are favourable to the passage of the excess silver ion into an interstitial position in which the jump time at room temperature is 10^{-10} sec and the diffusion constant, $1{\cdot}2\times10^{-5}$ cm^2 sec^{-1}. Grimley (1950 a) has, moreover, calculated a value of $-3{\cdot}1$ ev for the energy change accompanying the transfer of a silver ion from infinity to an interstitial position in the crystal.

We conclude that this mechanism (Mitchell 1957 a, b) is a possible one and that it results in the low-lying unoccupied energy level associated with the excess silver ion (Ag^+X, Ag_2S^+), being raised to the position of the shallow trapping levels provided by interstitial ions, 0·08 ev below B (see Simpson 1949, 1955).

In the two cases considered above, this leaves (1) a halogen atom adjacent to a halide *ion* with which it combines to form a negatively charged halogen molecule X_2^- at a site adjacent to a kink site occupied by a silver *ion* and (2) an adsorbed AgS diatomic molecule. Neither the interstitial silver ion nor either of these adsorbed molecules can trap an electron at room temperature, so that a clear explanation is provided for the absence of immediate recombination and for the unique position held in photographic sensitivity by silver halides with the rock-salt structure in which the lattice disorder is of the Frenkel type and the interstitial silver ions have a high mobility.

The photochemical process which has so far been discussed constitutes the primary stage in the formation of the latent image. This stage comprises the absorption of energy in the crystals and the liberation of conduction electrons and an equivalent number of interstitial silver ions. We have next to consider the combination of these interstitial silver ions with the electrons to form the latent image specks ; this we shall call the secondary stage in the formation of the latent image.

The latent image specks are formed by the combination of silver ions and electrons at sites adjacent to silver *ions* on kink and edge sites on the external surfaces and on jog and edge sites along internal edge dislocations. For the understanding of the mechanism which will be proposed, it is important to realize that these surface silver *ions* do not carry the effective localized charges which would be estimated on the assumption that the silver halides have properties closely approximating to those of ideal ionic crystals such as potassium chloride ; for this reason, *ion* has been printed in italic type in this paper whenever reference has been made to a surface *ion*.

The kink sites and jogs have half unit charges from the point of view of long-range electrostatic interaction with other charges. The depths of the potential energy wells which they furnish for the trapping of charges of opposite sign are, however, reduced by polarization ; this is an important factor in substances with partial homopolar binding, such as silver chloride and silver bromide, which have high dielectric constants. A silver ion has a much greater ionization potential than an alkali ion (7·54 ev compared with approximately 5 ev) ; a silver ion at a kink site or jog is therefore overlapped to a greater extent by the atomic p-orbitals of the neighbouring halide ions, with consequent reduction in the depth of the positive potential energy well which it furnishes for the trapping of an electron. A very approximate estimate of the interaction energy between a silver ion on a kink site or jog and a conduction electron may be made from the relationship $W=-13{\cdot}5Z^2/\epsilon^2$ ev ; with a value of $Z=+\frac{1}{2}$ and an average value for the dielectric constant ϵ of the environment between 8 and 13·3, W is found to lie between 0·05 and 0·02 ev. From such reasoning, it appears most unlikely that silver *ions* on surface kink sites and jogs can provide stable traps for conduction electrons at room temperature.

At low temperatures, the concentration of interstitial silver ions should be greater in the immediate neighbourhood of the surface and of edge dislocation lines than within perfect regions of the crystal ; the free energy can be reduced by the passage of a small fraction of the silver ions on kink sites and jogs into interstitial positions, leaving an excess of these sites occupied by halide ions. The surfaces and the edge dislocation lines therefore carry a negative charge which is compensated by a diffuse space charge of interstitial silver ions. The drift of these interstitial ions in electric fields contributes to the structure sensitive ionic conductivity at low temperatures.

7.4. *The Secondary Stage in the Formation of the Latent Image*

7.4.1. *Latent pre-image specks.*

It has been pointed out in §§ 7.2.4 and 7.3 that neither silver *ions* on surface kink sites or jogs nor adsorbed silver sulphide molecules provide effective traps for electrons at room temperature. The traps for the conduction electrons appear, as discussed in § 4.5, when the shallow positive potential energy well, due to a surface silver *ion*, is deepened by the presence of an interstitial silver ion in the immediate neighbourhood (Mitchell 1951 a, 1952). The conduction electron is then drawn into the space between the surface *ion* and the interstitial silver ion. It combines with the surface *ion* to form a silver atom and the local negative charge which would otherwise arise is neutralized by the interstitial silver ion. An unstable latent pre-image speck is thus formed which can be described as a silver atom adsorbed at a site adjacent to a silver *ion* on a surface kink site or jog.

We deduce from our interpretation of the results of Webb (1950 a) on the temperature dependence of low intensity reciprocity failure (§ 8.2) that the thermal dissociation energies of surface and internal latent pre-image specks are 0·77 and 0·65 ev respectively. Dissociation occurs through the passage of a silver ion into an interstitial position, an ionic process for which τ_0 in the equation

$$\tau = \tau_0 \exp(W/\boldsymbol{k}T)$$

would be 10^{-12} sec. With this value, the mean lifetimes of the latent pre-image specks at room temperature on the external and internal surfaces of the sensitivity centres are 24 and 0·2 sec respectively.

From this, it is clear that the formation of a more stable latent sub-image speck requires the combination of a second electron and interstitial silver ion with a latent pre-image speck within the total excitation time of the first electron. This is determined by the time which it spends in latent pre-image specks, its lifetime in the conduction band being negligible in comparison. If a second electron and interstitial silver ion are not made available by the absorption of a second quantum within this time, the first electron and interstitial silver ion recombine with the X_2^- or AgS molecules ; this restores the original situation and there is then no photochemical change. If a second quantum is absorbed while a latent pre-image speck from a previous event is still available, the positive hole can be trapped either by the X_2^- molecule to form a neutral X_2 molecule which can diffuse away from the surface and possibly react with groups of silver atoms suspended in the gelatin or else by a second silver sulphide molecule. In either case, a second conduction electron and interstitial silver ion become available for the formation of the latent sub-image speck.

The positive potential energy well, associated with a latent pre-image speck, is not sufficiently deep to trap a conduction electron at room temperature. The depth has again to be increased by the presence of an interstitial silver ion in the neighbourhood. The electron is then trapped and the negative charge is neutralized by the interstitial silver ion with the formation of a latent sub-image speck. With these mechanisms, no localized negative charge which can exert a long-range attraction for positive holes ever appears during the formation of latent pre- and sub-image specks. The dependence of the efficiency of the formation of the latent image during the early stages upon the concentration of available silver ions (see § 8.3) is also clearly explained.

The efficiency with which latent sub-image specks are formed depends on the mean lifetime of the latent pre-image specks at any particular temperature and on the mean time interval between the absorption of successive light quanta. Maerker (1954) has measured the critical inter-quantic times for four commercial Kodak emulsions and three experimental emulsions at the Kodak Research Laboratories in Rochester. Three of the commercial emulsions gave a value of 3 seconds (0·72 ev), and one a value of 4 seconds (0·725 ev). The experimental emulsion in the primitive chemically non-sensitized state gave a value of 4 seconds for the surface latent image which was increased to 6 (0·735 ev) and 10 (0·75 ev) seconds, respectively, by sulphur and reduction sensitization. The critical inter-quantic times for the internal latent image in the primitive, sulphur and reduction sensitized states were 5 (0·73 ev), 4 and 3 seconds, respectively. The values given in brackets are the dissociation energies of the latent pre-image specks which correspond with mean lifetimes equal to the critical interquantic times ; a vibration period of 10^{-12} sec has been used in the calculations.

From these observations, it is evident that the mean lifetimes of the latent pre-image specks are practically independent of chemical sensitization and have essentially the same values for the chemically sensitized commercial emulsions and for the surface and internal latent images in the primitive emulsion. The only conclusion which can be drawn from this is that the secondary process in the formation of the latent image is not influenced by the methods which are used for the sensitization of these particular commercial emulsions and is essentially the same as for both the surface and the internal latent images in the primitive emulsion. There is some inconsistency between the critical interquantic times measured by Maerker for the surface and the internal latent images and those which may be calculated from the observations of Webb (1950 a) on the temperature dependence of the low intensity reciprocity failure of the corresponding latent images, but this is perhaps not surprising in view of the probable variation in the detailed characteristics of the different emulsions.

All these observations support the new theory of photographic sensitivity and the discussion of the formation of latent pre- and sub-image specks during the secondary process which is being presented here. They are clearly inconsistent with the basic postulate of the Gurney–Mott theory of the formation of the surface latent image which was that the products of chemical sensitization provide deeper traps for electrons than any available in the chemically non-sensitized crystals.

7.4.2. *Surface and internal latent sub-image specks.*

Latent sub-image specks consist of pairs of silver atoms, formed by two successive events, at sites adjacent to silver *ions* on the external and internal surfaces of the sensitivity centres. The lifetimes of surface latent sub-image specks are of the order of days (Burton 1946, 1948 a, b, Burton and Berg 1946, Berg and Burton 1948) whereas those of latent pre-image specks are of the order of seconds. Latent sub-image specks have shorter lifetimes on the internal than on the external surfaces of the sensitivity centres.

The latent sub-image is formed by a distribution of latent sub-image specks which introduces a group of occupied energy levels in the forbidden band. The wavelength dependence of the Herschel effect (§ 9.2) shows that this group has a maximum density about 1·54 ev below B. The latent sub-image specks can therefore trap positive holes to form positively charged latent pre-image specks which then dissociate into interstitial silver ions and conduction electrons, and it is for this reason that a surface latent image is not formed in chemically non-sensitized crystals. Before any latent image can be formed, positive holes must be trapped at the surface so that halogen molecules can be formed which can escape. The surface latent sub-image specks provide deeper traps for positive holes than surface halide *ions*. Halogen molecules are therefore formed with low probability as long as latent sub-image specks are present on the external surface. Interstitial silver ions and electrons, chemically equivalent to the halogen molecules which are actually formed and escape, combine upon the internal surfaces of the sensitivity centres to build up the internal latent image.

A latent sub-image speck does not provide a stable trap for an electron at room temperature; the positive potential energy well due to the presence of the silver *ion* at the adjacent kink site or jog is not sufficiently deep, and the electron affinity of the pair of silver atoms cannot exceed about 1·5 ev.

7.4.3. *Chemical sensitization and the formation of the latent sub-image.*

The principal function of silver sulphide in photographic sensitivity is to provide far more numerous traps for positive holes than the latent sub-image specks and thus to protect these specks from regression. This function cannot be performed by adsorbed halogen acceptors, in which the highest occupied electronic energy levels lie below the top of the full band of the crystal. This explains why they do not sensitize emulsions for the formation of a surface latent image.

If atoms of gold were adsorbed in interstices and around the edges of the monolayers of silver sulphide at the sensitivity centres, they could combine with interstitial silver ions and conduction electrons; latent sub-image specks could then be formed by the absorption of single quanta in the microcrystals and reciprocity failure should disappear if development were carried to completion. The results of Maerker (1954) show that this situation is not realized in many commercial emulsions and that any groups of silver and gold atoms which are formed during chemical sensitization are unable to trap either interstitial silver ions or conduction electrons to influence the secondary processes in the formation of the latent image. In these emulsions, nuclei upon which interstitial silver ions and electrons can combine to form latent image specks are produced, not by chemical methods during sensitization, but, as latent sub-image specks, during the initial stages of exposure.

The situation may be different in fine grain experimental emulsions which have received the maximum possible combined sulphur and reduction sensitization. The nature of the resulting sensitizing layers in these emulsions has been discussed in §§ 5.4 and 5.6. Such emulsions fog rapidly on storage and it is possible that silver atoms which can combine with silver ions and electrons during exposure are co-adsorbed with the silver sulphide as discussed in § 5.4 (see also § 8.3). Even if this does not happen, the groups of silver atoms which are associated with the gelatin molecules will always combine with molecules of halogen liberated at the surfaces of the crystals and increase the efficiency with which the latent image is formed by reducing the amount of regression.

7.4.4. *The formation and properties of latent image specks.*

Photographic emulsions in which a latent sub-image has been produced by a hypersensitizing high intensity pre-exposure show neither a toe to the characteristic curve nor low intensity reciprocity failure (Berg 1948). From this we conclude that a latent sub-image speck is transformed into a latent image speck by a single photochemical event and that the latent image specks of minimum size consist of groups of three silver atoms.

A latent sub-image is converted into a latent image by prolonging the exposure. The latent sub-image specks act as combination centres for silver ions and electrons according to the following equation :

$$Ag_2 + Ag^+ + e^- \rightleftharpoons Ag_3{}^+ + e^- \rightarrow Ag_3.$$

The interstitial silver ion is adsorbed to the speck to charge it positively and the positive charge is then neutralized by the conduction electron. The latent sub-image specks are, however, not completely stable. They dissociate on prolonged storage and the resulting conduction electrons and interstitial silver ions combine with other latent sub-image specks by this process to increase their size to that of latent image specks. The thermally activated condensation process occurs more rapidly on the internal than on the external surfaces where the latent sub-image specks are stabilized by adsorption to silver sulphide and by the presence of gelatin. It may be accelerated by post exposure. Some latent sub-image specks then trap positive holes or lose electrons and dissociate ; the interstitial silver ions and the electrons combine with other latent sub-image specks to form latent image specks.

So far, we have been concerned with silver atoms (latent pre-image specks), pairs of silver atoms (latent sub-image specks) and groups of three silver atoms, probably in a triangular arrangement (latent image specks), adsorbed adjacent to surface sites occupied by silver *ions*. These groups increase in size by a process in which an electron is trapped while the interstitial silver ion, which neutralizes the charge, is present in the immediate neighbourhood or already adsorbed to the group. In these circumstances, a local negative charge which could attract and combine with a positive hole never appears and the difficulty mentioned in § 4.5 in connection with the theories of Gurney and Mott and of Seitz does not arise. We must, however, explain why the latent image specks themselves do not trap positive holes because it does not seem likely that the small effective charges of the silver *ions* on the kink and edge sites where they are formed would ensure either the repulsion of positive holes or the trapping of electrons. Positive holes would certainly be trapped if the localized energy levels occupied by the valency electrons of the latent image specks were distributed in the forbidden band above C. The resulting positively charged specks would attract and combine with electrons, so that the latent image specks would serve merely as recombination centres for positive holes and electrons and the further separation of silver atoms would proceed only with very low quantum efficiency (see § 7.1.3).

A related observation which also requires explanation is that, as noted in § 7.1.1, the induction period for development decreases substantially when latent image specks are formed from latent sub-image specks. It does not seem possible to account for this if the only change is an increase in the size of the speck by one silver atom.

These problems are both resolved if the latent image specks, which are groups of three silver atoms, adsorb adjacent surface silver *ions* and form positively charged groups of four tetrahedrally disposed silver atoms, and vacant silver ion lattice sites, which remain in the neighbourhood to maintain electrical neutrality.

The positive charge lowers the occupied energy levels in the latent image specks below the top of the full band, and the groups repel and do not trap positive holes. They provide deep traps for electrons, and the positive charge is then restored to the uncharged group by the adsorption of a further silver ion from an adjacent surface site with the formation of a second vacant silver ion lattice site. The positive holes are trapped not by these vacant silver *ion* lattice sites but by other localized occupied surface levels in the forbidden band, and interstitial silver ions drift from the trapping sites to the neighbourhood of the latent image specks to neutralize the space charge due to the excess vacant silver ion lattice sites.

This concentration mechanism, which accounts for the formation of a few discrete latent image specks, cannot operate until a group of three or more silver atoms has been formed. The heat of adsorption of a silver *ion* from a surface lattice site is then sufficient to ensure that the speck is positively charged, on the average, when it is in equilibrium with the crystal at room temperature.

The smaller latent pre- and sub-image specks cannot adsorb silver *ions* from adjacent surface sites in this way and depend upon the presence of interstitial silver ions for the trapping of electrons and their consequent growth. There are many equivalent shallow potential energy wells associated with surface silver *ions* and concentration effects are not in evidence at these stages in the formation of the latent image. This situation changes when once latent image specks, which adsorb silver ions and become positively charged, have been formed. The presence of the positive charge lowers both the unoccupied and occupied energy levels in the specks with respect to B and C. These levels are raised when the positive charge is neutralized by the trapping of an electron and lowered again by the adsorption of a further silver ion. The positively charged latent image specks thus repel positive holes and provide deep traps for electrons. The silver atoms are concentrated in a manner very similar to that envisaged by Gurney and Mott (1938) in their theory of photolysis, with the fundamentally important difference that the adsorption of the silver ion precedes the trapping of the electron by the neutral speck.

7.5. *Conclusion*

The solutions of the difficulties in the theory of the formation of the latent image which were raised in § 7.1.3 will now be summarized. Molecules and atoms of chemical sensitizers do not trap conduction electrons as has usually been assumed in the past. They trap positive holes, and recombination between the resulting positively charged centres and the conduction electrons is prevented by the dissociation of the centres and the passage of the excess surface silver ions into interstitial positions. This removes the unoccupied energy levels near the top of the full band and replaces them with unoccupied energy levels near the bottom of the conduction band associated with the interstitial silver ions. At room temperature, the absorption of energy in the crystals thus leads to the liberation of conduction electrons and the appearance of an equivalent number of interstitial silver ions.

Silver *ions* at surface sites do not trap conduction electrons at room temperature. Latent pre-image specks, which consist of silver atoms adsorbed at sites adjacent to surface silver *ions*, are formed only when the depths of the traps are increased by the presence of interstitial ions in the immediate neighbourhood. The conduction electron and the interstitial silver ion then combine at the site to form an adsorbed silver atom, or a latent pre-image speck, which has only a transitory existence at room temperature. The occurrence of a second event of the same nature transforms a latent pre-image speck into a more stable latent sub-image speck which consists of a pair of silver atoms adsorbed at a site adjacent to a surface silver *ion*. With these mechanisms, a localized negative charge which can attract a positive hole never appears during the early stages in the formation of the latent image. The importance of the presence of excess silver ions, a fraction of which may be present as a diffuse space charge of interstitial silver ions, for the formation of latent pre- and sub-image specks also becomes clear, and an explanation of the dependence of low intensity reciprocity failure (see §§ 8.2 and 8.3) upon the concentration of silver ions in the system follows.

Discrete groups of silver atoms are formed because of the instability of latent pre-image specks. If these are not converted into latent sub-image specks by the occurrence of a second event during their limited lifetimes, they dissociate, and the conduction electrons and interstitial silver ions combine with latent image specks to increase their size. In these circumstances, a few large latent image specks are formed.

The explanation of the fact that latent image specks and larger visible particles of silver do not provide recombination centres for positive holes and conduction electrons is based upon the postulate that, at room temperature, groups of three or more silver atoms adsorb adjacent surface silver *ions* and thus become positively charged. The compensating negative charge is provided by vacant silver ion lattice sites in the neighbourhood. This lowers the highest occupied energy levels in the specks below the top of the full band, so that positive holes are not trapped by the specks.

The trapping of a conduction electron produces an uncharged speck which does not exert any long-range attraction for a positive hole ; neither does the vacant silver ion lattice site, which is, in any case, not immediately adjacent to the speck, trap a positive hole at room temperature. The positive charge is restored to the latent image speck by the adsorption of a further surface silver *ion*, and a second vacant silver ion lattice site is created. The positive hole is trapped at some other surface site, and the interstitial silver ion which is made available diffuses to the neighbourhood of the latent image speck and combines with one of the vacant silver ion lattice sites to remove the potential difference created by the trapping of the electron and the positive hole at separated sites. The detailed mechanisms which have been proposed are thus in complete accord with the Gurney–Mott principle governing photochemical changes in the silver halides (see Mitchell 1957 a, b).

§ 8. RECIPROCITY FAILURE

8.1. *Introduction*

The density produced after a given exposure (intensity × time) by developers of low activity, or by incomplete development with active developers, is, for many photographic materials, not independent of the intensity at which the exposure is made. The curve obtained by plotting the logarithm of the exposure required to produce a given density against the logarithm of the intensity, has a minimum corresponding with an optimum intensity of exposure. If the Bunsen–Roscoe reciprocity law for photochemical changes were obeyed, the exposure would be independent of the intensity, and the relationship would be represented by a straight line parallel to the intensity axis. With practically all emulsions, longer exposures are required to produce the same density at low and at high intensities and this is described as low and high intensity failure of the reciprocity law. This aspect of photographic sensitivity has always to be taken into account in the standardization of emulsions for photometric purposes.

Ten years ago, it appeared that the experimental work of Evans and Hirschlaff (1939), Webb and Evans (1938 a, b), Webb (1942) and Berg (1939, 1940) on the temperature dependence of low and high intensity reciprocity failure had provided strong support for the Gurney–Mott theory of the formation of the latent image (Berg 1948). Since then it has become evident that the position requires careful re-examination. Webb and Evans (1938 b) assumed that low intensity reciprocity failure arose from a regression of the small latent image specks in the initial stages of their formation. The disintegration of the specks was thought to be the result of thermal agitation, and it was supposed that the ejection of an electron from the speck was followed by the diffusion away of a silver ion. Gurney and Mott (1938) had themselves proposed a theory of low intensity failure, in which they postulated that the electron was ejected from the sensitivity speck (assumed to be an uncharged speck of silver sulphide) as a result of thermal agitation before it had combined with a silver ion, so that the latent image did not form until the concentration of electrons in the crystal reached a certain critical value. According to either theory, low intensity reciprocity failure would be diminished by lowering the temperature, which would increase the lifetimes of the electrons in the traps.

Webb (1950 a, b) has since given a more detailed discussion of low intensity reciprocity failure and, from experimental work on the displacement of the low intensity branch of the reciprocity failure curve as a function of temperature, [illegible] has determined values of 0·77 ev and 0·65 ev for the activation energies of the temperature-dependent processes responsible for low intensity reciprocity failure in the formation of the surface and internal latent images. From his discussion, it appears that Webb assumes that this is the energy required for the thermal ejection of an electron from a latent sub-image speck and that, allowing for the fact that the thermal activation energy for the ejection of an electron from a centre is sometimes about one half of that required for optical ionization (Mott and Gurney 1948), it is to be compared with the energy of the quantum (1·54 ev) required to produce the Herschel effect which has a maximum sensitivity at about 8000 Å. It is difficult to reconcile this discussion with the observations of Burton (1946, 1948 a, b, see also Burton and Berg 1946), which suggest that latent sub-image specks have lifetimes of hours or days instead of the lifetimes of seconds which would be required for the interpretation of low intensity failure. A similar conclusion is reached from the studies which use the Herschel effect to follow the transformation of a latent sub-image into a latent image during the storage of an exposed emulsion (§ 9.2).

As the latent sub-image speck, according to the analyses of Katz (1949, 1950) and Webb (1950 a, b), consists of only two adjacent silver atoms formed by the absorption of two quanta within a critical time τ, the Gurney–Mott theory would seem to imply that the intermediate stage in the formation of the latent image leading to low intensity reciprocity failure involved silver atoms which trapped electrons and then attracted distant interstitial silver ions; all the evidence suggests that silver atoms in crystals of silver halides could not trap electrons at room temperature.

There is a further difficulty with the treatment of low intensity reciprocity failure proposed by Katz and by Webb. They both assume that the thermally activated disintegration process is the ejection of an electron from a trap. They find, however, that, in order to obtain a value for the average lifetime of the unstable centre which is in agreement with the experimentally determined values of τ, the time within which two quanta have to be absorbed to produce a latent sub-image speck (see Maerker 1954), they are compelled to use a period τ_0 of 10^{-12} sec in conjunction with the experimentally determined activation energy in the following equation:

$$\tau = \tau_0 \exp(+W/kT).$$

10^{-12} sec is the period of vibration of the ions of the lattice, and a value nearer 10^{-8} sec is usually found for processes involving the thermal excitation of bound electrons. These difficulties are inherent in all the models which have so far been used for the interpretation of low intensity reciprocity failure on the basis of the Gurney–Mott theory of the formation of the latent image, which requires the trapping of conduction electrons to be followed by the neutralization of the resulting negatively charged traps by the motion of interstitial silver ions.

Difficulties also arise in the discussion of high intensity failure in terms of this theory. On the basis of the Gurney–Mott theory, it was deduced that a second electron could not be trapped at a site until the negative charge due to the trapping of the first had been neutralized by an interstitial silver ion, a process which would require a definite time on account of the low mobility of the interstitial silver ions. The result would be that a dispersed latent image which could not initiate development efficiently would be formed at high intensities. If this were the actual process the activation energy derived from the temperature dependence of the displacement of the high intensity branch of the reciprocity curve should be that for the motion of either interstitial silver ions (0·15 ev) or vacant silver ion lattice sites (0·36 ev). The value found by Webb (1942) of 0·69 ev does not correspond with either of these values or even with the formation energy of a Frenkel defect (1·27 ev). It is, in fact, near the activation energy of the process responsible for low intensity failure. The conclusion that the Gurney–Mott theory of the formation of the latent image cannot be applied to the interpretation of high intensity failure, as has previously been done by Mott (1941), Webb (1942) and Berg (1940, 1948), appears to be inescapable.

8.2. *The Interpretation of Reciprocity Failure*

The difficulties outlined in the previous section disappear when it is recognized that the same thermally dependent process is involved in both low and high intensity reciprocity failure. This process is the dissociation of the latent pre-image speck, which appears as an essential intermediate stage in the formation of a latent sub-image speck, with the loss of a silver ion into an interstitial position (Mitchell 1957 a, b). An ionic process is involved, for which the vibration period is about 10^{-12} sec, and the activation energy is greater than that for the diffusion of an interstitial silver ion but not as great as that for the formation of a Frenkel defect. At extremely low intensities, the photochemical process is inefficient because electrons and interstitial silver ions recombine with X_2^- groups or AgS molecules before the occurrence of second events which allow halogen molecules to escape from the surface and transform latent pre-image specks into latent sub-image specks.

During exposures at low intensities, some latent pre-image specks are required to survive for the interquantic time—the mean time between the absorption of two successive quanta by the crystals—in order that latent sub-image may be formed, and regression avoided. The fraction of the latent pre-image specks with lifetimes greater than the interquantic time increases, and low intensity reciprocity failure consequently decreases, as the temperature is lowered. The latent pre-image specks which have the longest lifetimes, and therefore the greatest dissociation energies, are the most important ones for low intensity exposures. We conclude that the activation energies required for their dissociation on the external and internal surfaces of the sensitivity centres are the values 0·77 and 0·65 ev determined by Webb (1950 b) from his observations on the displacement of the low intensity branches of the reciprocity failure curves with changing temperature.

When once latent sub-image specks have been formed during low intensity exposures, further interstitial silver ions and conduction electrons combine with them to form latent image specks and the condensation mechanism discussed in § 7.4.4 then becomes operative. Latent pre-image specks formed at this stage are more likely to dissociate than to be stabilized by a second event, so that latent sub-image specks which can provide further condensation nuclei are not formed and the latent image consists of a few large latent image specks. This account of the reasons for low intensity failure, together with the interpretation of its dependence on the concentration of silver ions in the system given in § 7.5, is consistent with all the published observations on the subject.

The quantum efficiency increases with increasing intensity because of reduced regression, but the developability of the latent image which is formed decreases, because more latent sub-image is formed by the combination of interstitial silver ions and electrons with latent pre-image specks. Eventually a point is reached at which all the latent pre-image specks become latent sub-image specks and no latent image specks are produced during exposure. Thereafter, high intensity reciprocity failure disappears, as was first observed by Berg (1940), but the overall process appears to be inefficient because of the long induction period required for the initiation of development by the latent sub-image.

When exposures are made at high intensities, some latent pre-image specks are required to dissociate within the interquantic time, in order that the interstitial silver ions and electrons may combine with latent sub-image specks to increase their size to that of latent image specks. It is therefore clear that the latent pre-image specks which have the shortest lifetimes and smallest dissociation energies are now the important ones. The fraction of latent pre-image specks with lifetimes less than the interquantic time is increased and high-intensity reciprocity failure correspondingly reduced by raising the temperature.

As mentioned in the previous section, Webb (1942) deduced a value of 0·69 ev for the activation energy of the temperature dependent process involved in high-intensity reciprocity failure from his studies of the displacement of the high-intensity branch of the reciprocity failure curve with changing temperature. We interpret this as the minimum dissociation energy of a latent pre-image speck. That this value is less than the value, 0·77 ev, which Webb (1950 a) obtained from measurements on the low intensity branch follows directly from the interpretation of reciprocity failure which has been given above. It is surprising that the difference is not greater. A value of 0·4 for the activation energy would be required to explain the occurrence of the Berg bend over point (see Berg 1940) at an exposure time of 10^{-5} second. The mechanism proposed here provides a clear explanation of the fact that the reciprocity failure curves are displaced along the intensity axis without much change of shape with changing temperature (see Berg 1948). This behaviour, which was difficult to understand on the basis of the Gurney–Mott theory, is fully consistent with the new theory in which the occurrence of both low and high intensity failure is determined by the same temperature-dependent variable.

Our conclusion is, therefore, that the minimum in the reciprocity failure curve, obtained by plotting the exposure required to produce a given density when development is not carried to completion, arises from the balancing of two factors.

The first is the inefficiency of the photochemical process at low intensities, due to regression which decreases as the intensity is increased. The second is the inefficient initiation of development by a latent sub-image, which reduces the density produced by incomplete development to an ever-increasing extent with increasing intensity, because a steadily increasing fraction of the latent image appears in the form of latent sub-image specks. High-intensity reciprocity failure disappears, according to this model, when the intensity reaches a value at which the latent image consists almost entirely of latent sub-image specks.

The detailed mathematical treatments of low intensity reciprocity failure by Katz (1949, 1950) and by Webb (1950 a, b) may be applied to the model introduced here.

8.3. *The Reduction of Reciprocity Failure*

The experimental work of Clark and Mitchell (1956) on chemically sensitized large crystals of silver halides, which show marked low intensity reciprocity failure, established that the only effective method for reducing low intensity failure was to bathe the sensitized crystals in solutions of silver nitrate at a *pAg* between 5·4 and 3·0 and to dry them before exposure. In these experiments, it was found that low intensity failure always increased as the availability of mobile silver ions at the surface of the crystal was reduced, whether by the incorporation of cadmium or lead ions in mixed silver halide crystals (see also Fujisawa and Mizuki 1951) or by bathing sensitized crystals in solutions of (i) alkali bromides, (ii) strongly adsorbed antifoggants, or (iii) weakly adsorbed stabilizers, before exposure. This emphasized the important role of excess silver ions in low intensity reciprocity failure.

Low intensity failure is also reduced by prolonging digestion during chemical sensitization (Webb and Berg 1954, Mizuki and Fujisawa 1956). The nature of the resulting sensitization has been considered in detail in § 5.4. Latent sub-image specks could be produced by single photochemical events if silver atoms were co-adsorbed with silver sulphide monolayers (see § 7.4.1) ; reciprocity failure would then be greatly decreased provided that development was carried to completion. The results of Maerker (1954) show, however, that this cannot be the important mechanism. The groups of silver atoms which are associated with adsorbed gelatin molecules probably react with halogen which is liberated at the surface and prevent low intensity reciprocity failure of the type observed by Berg (1948).

High intensity reciprocity failure is essentially a development phenomenon, connected with the long induction period required for the development of a latent sub-image. It disappears when development is carried to completion or when highly active developers containing development accelerators are employed ; these matters are considered in § 11, together with the reasons for the reduction in high-intensity reciprocity failure which results from gold sensitization and latensification.

The reduction of low intensity failure by a uniform, high intensity, hypersensitizing pre-exposure and that of high intensity failure by a uniform, low intensity, latensifying post exposure were fully described by Berg (1948) in the previous review, and nothing more need be added here. Hautot and Sauvenier (1951 a) have reported further experimental work on latensification by low intensity post exposure.

§ 9. LATENT IMAGE EFFECTS

9.1. *The Formation of the Latent Image at Low Temperatures*

The treatment of the formation and nature of the latent image, which has been given in § 7, can be applied to the interpretation of all the latent image effects and to the discussion of the formation of the latent image as a result of exposure at low temperatures. These subjects were considered at length by Berg (1948) and a further detailed review will not be given in this paper. The way in which the observations can be interpreted in terms of the new theory will be evident from the previous accounts.

The exposure of photographic emulsions at low temperatures, at which interstitial silver ions are immobile, is accompanied by luminescence (Meidinger 1940, 1944). The relevance of this phenomenon to the photochemical process does, however, not yet appear to have been fully established (Farnell 1952, Biltz 1952). Biltz (1949, 1952) made the important observation that the surface sensitivity is decreased to a greater extent than the internal sensitivity at low temperatures. This implies that chemical sensitization (assumed to be due mainly to silver sulphide) becomes ineffective. He observed a new form of high-intensity reciprocity failure at low temperatures and suggested a correlation between this phenomenon, the loss of surface sensitivity and the luminescence. This is consistent with the theory of the formation of the latent image presented in § 7.

Other work on the formation of the latent image as a result of exposures at low temperatures has been carried out by Vassy and Vassy (1948) and Falla (1952, 1953 a, b, 1954, 1955 a, b). Falla (1955 b) has discussed the general validity of the conclusions reached by Berg (1943) as a result of his work at low temperatures.

9.2. *The Herschel and Debot Effects*

In this section we shall review the work on the Herschel effect which has been the subject of much research during the past ten years, and describe the complementary Debot effect (see Berg 1947, Hautot and Sauvenier 1949, Falla 1952). The Herschel effect is a reduction in the density produced by surface development, which occurs when emulsions which have been exposed to actinic light are later exposed to red or infra-red light to which they are not normally sensitive. It is due to a redistribution of the latent image, in which the surface latent image is weakened and the internal latent image built up.

For the investigation of the Debot effect (Debot 1941, 1942, Hautot and Sauvenier 1949), an emulsion is exposed to produce an internal latent image and is then treated with an oxidizing solution which destroys the surface latent image without affecting the internal image. When the emulsion, which has been prepared in this way, is exposed to red or infra-red radiation, it is found that a surface latent image is formed, due to the transference of silver from internal surfaces to the external surface, which may be developed with a surface developer.

Taken together, these two effects imply that slow changes must occur in the distribution of the silver atoms of the latent image between the surface and the interior of the crystals during prolonged storage and that these changes are accelerated by irradiation with red or infra-red light. This is consistent with the instability of a distribution of silver atoms on a silver bromide crystal, which was established by Evans, Hedges and Mitchell (1955).

The recent investigations of the Herschel effect have been particularly valuable in providing evidence of the nature and stability of the latent image. It was first appreciated by Urbach and Wolinski (1938) that the Herschel effect could be used to detect those changes occurring during exposure or immediately afterwards which could not be established by more direct methods. It involves the ejection of an electron from the latent sub-image speck or from an uncharged latent image speck, and the dissociation of the remaining positively charged speck. Information is thus made available on the stability of uncharged latent image specks against the loss of electrons by infra-red irradiation and thermal excitation and on the stability of the positively charged specks which are their immediate precursors.

Kornfeld (1949) confirmed the observations of Lescynski (1926) and Meiklyar (1946) that the Herschel effect is much greater for the latent sub-image produced by a high intensity exposure than for the latent image produced by a low intensity exposure. This is fully consistent with the proposal made in § 7 that latent sub-image specks are neutral groups of two silver atoms whereas latent image specks are normally positively charged groups of at least four silver atoms. The steady decrease in the Herschel effect which occurs during the storage of an exposed emulsion over a long period of time at room temperature or at higher temperature, which had been noted by Hautot and Falla (1942), has also been confirmed by Kornfeld (1949).

Kornfeld and James (1943) concluded that the loss of one silver atom from a surface latent sub-image speck was responsible for the loss of the developability of one grain. Kornfeld (1949) also confirmed the observation of Kornfeld and James (1943) that the Herschel effect did not change during the induction period in the development of the surface latent image in a fine grain positive emulsion with a standard metol hydroquinone developer. This shows that the characteristics of the latent image are not changed during the induction period, and is in agreement with the interpretation of the induction period which will be given in § 11.

Vanselow, Quirk and James (1953) have studied the effect of oxygen and moisture on the Herschel effect and found a complicated pattern of behaviour. Vanselow and James (1950) studied the influence of degree of development on the Herschel effect and concluded that the absolute number of grains rendered non-developable by the infra-red exposure was independent of the development time and the developer used. In no instance was the Herschel effect eliminated by prolonged development. The loss of developability is caused by the absorption of one quantum per centre and therefore by the loss of one silver atom per centre, although the remaining speck may subsequently disintegrate (Kornfeld and James 1943, James and Vanselow 1949).

James, Vanselow and Quirk (1948) and Fortmiller, James, Quirk and Vanselow (1950) have used the Herschel effect to follow the changes in latent sub-image specks due to latensification. They found that gold latensification eliminated the Herschel effect and indeed conferred red sensitivity upon the exposed film. This effect has also been observed by Hautot and Sauvenier (1953 a, 1957 a). This is consistent with the interpretation of the effect of gold sensitization (Burrow and Mitchell 1954, Mitchell 1957 a, b) which will be summarized in § 11. Latent sub-image specks containing gold atoms can adsorb silver ions and become positively charged. This lowers the unoccupied and occupied energy levels in the specks relative to the levels of the silver halide crystal. Higher-energy quanta have then to be absorbed to transfer electrons from occupied levels of the specks to the conduction band, and the transference of electrons from the full band to the unoccupied levels of the specks can be caused by the absorption of low energy infra-red quanta.

Fortmiller, James, Quirk and Vanselow (1950) attempted to restore the developability of the grains which had been lost as a result of an infra-red exposure by latensification (see § 11), in order to establish whether or not a stable non-developable speck, which would have to consist of a single silver atom, existed. They had already demonstrated that a latent sub-image speck, which consists of a pair of silver atoms, can initiate development, by establishing that latensification of a latent sub-image by post exposure to low intensity light is not obtained when development is carried to completion with energetic developers (James and Vanselow 1949). They found that neither low intensity post exposure nor gold latensification was effective, even though gold latensification completely stabilizes latent sub-image specks. James and Vanselow concluded that latensification will not make developable any significant number of centres which were originally smaller by one unit than the critical size necessary for initiating development. These important observations show that development can be initiated by a latent sub-image speck consisting of a pair of silver atoms, but not by a single silver atom even if such an atom should be adsorbed in the silver sulphide monolayer and stabilized by the adsorbed layer of gelatin. This is consistent with the discussion of chemical sensitization given in § 5.4.

It is, moreover, well known that the Herschel effect is decreased or increased by raising or lowering the concentration of silver ions in the medium. This is in agreement with its dependence upon the instability of positively charged latent pre-image specks which must dissociate to avoid recombination with the conduction electrons.

Taken as a whole, the recent observations on the Herschel effect thus provide strong evidence to support the general approach to latent image theory which has been outlined in §§ 5, 7 and 8. They emphasize that it is essential to discuss the properties of the smallest possible groups of silver atoms, the latent pre- and sub-image specks, in connection with the formation of the latent image, and that useful conclusions cannot be drawn from arguments based upon the introduction of the physical properties of massive silver.

§ 10. SOLARIZATION AND PHOTOLYSIS

10.1. *Solarization*

Solarization is the term used to describe a reversal phenomenon arising from a steady decrease in the density produced by surface development which occurs when a time-scale exposure is increased beyond a certain limit. Solarization begins at greater exposures when the digestion of the emulsion during chemical sensitization is prolonged. It is largely prevented by bathing photographic materials in solutions of halogen acceptors and drying them before exposure.

The internal latent image does not normally show solarization and a reversed image is not usually produced by internal development.

The theoretical treatment of solarization has not changed since the review of Berg (1948), and the same alternative theories continue to be revived and discussed from time to time (Trivelli 1930, Hautot and Debot 1948, Hautot 1948, Belliot 1951, 1952). The experimental work of Hedges and Mitchell (1953 b), Evans and Mitchell (1955) and Clark and Mitchell (1956) has provided strong support for the rehalogenation theory. They concluded that the adsorbed silver sulphide which partially protects the surface latent image from regression is progressively exhausted during exposure, due largely to its reaction with the halogen, which is chemically equivalent to the silver which separates to form the internal latent image. When no silver sulphide remains, the surface latent image is itself attacked and the density produced by surface development then decreases. The silver sulphide may be protected from reaction with the liberated halogen by the introduction of a suitable halogen acceptor (Evans, Hedges and Mitchell 1955, Sutherns and Loening 1956 a, b), and this eliminates solarization. The internal latent image, which is formed on the adjacent internal surfaces of the sensitivity centres, continues to build up with increasing exposure and passes continuously into a visible image of colloidal silver.

The utility of the concept of the sensitivity centre introduced by Mitchell (1951 a, 1952, 1953 b, 1954), and of the approach to the subject of photographic sensitivity to which it leads, has been demonstrated in a particularly clear manner by the microscopic observations on solarization (Hedges and Mitchell 1953 a, b, Clark and Mitchell 1956, Mitchell 1956, 1957 a, b), which have established that the surface latent image is first solarized at surface sites associated with crystal imperfections within which internal latent image is formed.

Oh-Yama (1947 a, b, 1956) has shown that solarization begins after shorter exposures if the photographic material is subjected to pressure before exposure. This effect probably arises from an increase in the internal disorder of the crystals which leads to the formation of more internal latent image (Seitz 1956), with a corresponding increase in the amount of halogen liberated at the surface.

Sauvenier (1951, 1954) has demonstrated the solarization of the surface image formed by infra-red radiation ; this is of interest in connection with the mechanism of optical sensitization (see §§ 6.2 and 7.2.5).

10.2. *The Photolysis of Silver Halides*

10.2.1. *Introduction.*

Silver separates at an exceedingly slow initial rate when large single crystals of silver chloride and silver bromide, which have been grown from the melt and handled and annealed with extreme care to minimize the number of internal crystal imperfections, are exposed to wavelengths near the edge of the absorption band (Stasiw and Teltow 1941, Stasiw 1951, Clark and Mitchell 1956). It has to be concluded that the positive holes and electrons liberated by the absorption of energy in the crystals recombine in these circumstances.

The crystals may be sensitized either :

(1) by plastic deformation (Hilsch and Pohl 1930, Haynes and Shockley 1948, Pick and Pohl 1951, Hedges and Mitchell 1953 a, Mitchell 1957 c), or

(2) by the incorporation of divalent anions (Stasiw and Teltow 1941, Schöne 1953, Volke 1954, Stasiw 1954, 1955), or

(3) by the incorporation of cuprous ions (Clark and Mitchell 1956).

In imperfect and mechanically deformed crystals, silver separates along internal dislocation lines (Hedges and Mitchell 1953 a, b, Mitchell 1957 c). In cases (1) and (2) the same optical absorption band, due to colloidal silver, which was first studied by Hilsch and Pohl (1930), appears during exposure, but the crystals containing divalent anions have a greater sensitivity than mechanically deformed crystals of pure silver halide. According to Stasiw (1954), the colloidal silver specks formed by the absorption of energy in the impurity absorption band are distributed uniformly through the sensitized crystals, instead of being confined to within 40 μ of the surface as is the case with plastically deformed or structurally imperfect crystals of pure silver halides (Clark and Mitchell 1956).

In the third method, the cuprous ions appear to sensitize the crystals even more efficiently than divalent anions.

10.2.2. *Experimental work on photolysis in emulsions.*

Meidinger (1949 a, b) has measured the quantum yield for the production of silver in photographic emulsions and found a value of 0·45 at 3660 Å. The quantum yield is increased by raising the temperature at which the exposure is made, and by the addition of halogen acceptors. At 100°c, a quantum efficiency of 0·53 was found in the absence of a halogen acceptor. This was raised to 0·92 when the emulsion was exposed in a 1% solution of sodium nitrite. The experimental work of Eggert, Meidinger and Arens (1948), on the production of silver during the exposure of dye-sensitized emulsions in the absorption band of the dyestuff, has already been mentioned in § 6.1. Pouradier and Chateau (1954) have established that iodine, and not bromine, is liberated during prolonged exposures of silver bromo-iodide emulsions.

10.2.3. *Experimental work on photolysis in large crystals.*

The high quantum yields observed for the production of silver in photographic emulsions are of great interest, but the detailed interpretation of experimental results presents many difficulties on account of the inadequately defined nature of the variables involved. For this reason much experimental work has been done on the liberation of halogen and the separation of silver during the exposure of large single crystals and thin sheet crystals of silver halides.

(*a*) *The quantum yield for the production of halogen.* The vacuum photolysis of silver chloride and silver bromide has been studied by Luckey (1953, 1955), who followed the liberation of the halogen from evaporated layers and thin sheet crystals by pressure measurements with a Langmuir gauge and a platinum–iridium Pirani gauge. At 3650 Å, Luckey found quantum yields for bromine production of between 0·05 and 0·12 at 27°c. He observed that the quantum yield was not affected by incorporating 1% of divalent lead or cadmium ions in the crystals, and concluded that positive holes are not trapped by vacant cation lattice sites at room temperature. The quantum yield decreased as the temperature was increased. The incorporation of cupric ions sensitized the crystals at temperatures above 100°c, but did not give any significant effect on the quantum yield at room temperature. The quantum yields increased as the wavelength was decreased, from values of about 0·05 at 4360 Å to values near unity at 2800 Å. Luckey showed that the quantum yields for halogen production in thick crystals were governed by the relation $\phi = Q[1-\exp(-ka)]$, where Q is the surface efficiency with a value between 0·5 and 1, k is the absorption constant, and a is a parameter with values of 0·4 and 0·24 at room temperature for silver bromide and silver chloride respectively. This is equivalent to stating that light quanta have to be absorbed within a fraction of a micron of the surface of the crystal, to be effective in causing the liberation of halogen. The detailed theoretical analysis of this important work which is being undertaken by Duboc has not yet been published.

(*b*) *The quantum yield for the production of silver*. The first quantitative measurements of the quantum efficiency for the production of silver by silver halide crystals during the exposure were made by Hilsch and Pohl (1930), who estimated the amount of silver liberated from measurements of the area under the characteristic ‘latent-image’ optical absorption band (§ 2.9). They found a value of about 0·4 at a wavelength of 4050 Å. The crystals owed this relatively high sensitivity to previous plastic deformation (Pick and Pohl 1951). In more perfect silver bromide crystals, Stasiw and Teltow (1941), using the same method, found a value for the quantum efficiency of 0·02 at 4360 Å. They showed that well-annealed, large, single silver chloride crystals were very insensitive and that visible specks of silver separated during exposure only upon the external surfaces. These observations have been qualitatively confirmed by the microscopic observations of Hedges and Mitchell (1953 a, b) and Clark and Mitchell (1956), who also found that visible specks of silver formed only upon surfaces of silver chloride and silver bromide crystals which were free from internal structural imperfections. They demonstrated for the first time that silver separates along the individual dislocation lines in plastically deformed crystals, and thus accounted for the enhanced internal sensitivity of such crystals.

Stasiw and his co-workers have been responsible for many investigations on the sensitization of crystals of silver chloride and silver bromide by the incorporation of divalent anions (see § 5.4). The solubility of silver sulphide, silver selenide and silver telluride in silver chloride and silver bromide is extremely small at room temperature, but considerable at temperatures near the melting point. Mixed crystals containing 0·01–0·02 mole per cent of the impurity are therefore produced by melting the compounds together, crystallizing, annealing for long periods near the melting point to produce a homogeneous distribution, and then quenching rapidly to room temperature to preserve the distribution. Specimens for measurements on optical absorption and sensitized photolysis are then pressed out at a temperature near the melting point and quenched to room temperature. They are then immediately cooled to liquid-air temperature for the optical measurements. In these circumstances, well-defined impurity absorption bands, which are attributed to the excitation of the isolated divalent anions, have been observed at the following wavelengths :

	S^{2-}	Se^{2-}	Te^{2-}
AgCl	4550 Å	4730 Å	
AgBr	4900 Å	5120 Å	5420 Å

These bands are accompanied by further absorption at shorter wavelengths, which is assumed to cause the ionization of the impurity centres (Volke 1954, 1956, Seifert and Stasiw 1955, Seifert, Stasiw and Volke 1954). This absorption disappears when the crystals are kept for a period at room temperature before being cooled to liquid air temperature. Stasiw (1950, 1951, 1955) attributes this behaviour to the formation of association complexes between the divalent anions, which represent excess negative charges, and interstitial silver ions and vacant halide ion lattice sites, which are positively charged defects of the lattice. The absorption due to these association complexes is assumed to lie within the absorption band of the silver halide. Stasiw has emphasized that sensitization is observed only when the impurity is molecularly dispersed within the silver halide. Storage at room temperature results in the slow segregation of the impurity, and the sensitization of the photolysis then disappears. It does not, therefore, seem possible to apply the theory developed by Stasiw (1954, 1955) directly to the discussion of the detailed nature of sulphur sensitization in photographic emulsions, which is quite stable and shows no such slow disappearance on storage. Mitchell (1957 d) has pointed out that the silver sulphide which is effective in sensitizing silver halide crystals for the formation of the surface latent image is present in the form of a monomolecular layer adsorbed on the surfaces of the crystals, and is not molecularly dispersed within the crystals (see § 5.4).

Stasiw and his co-workers have studied the changes in the optical absorption which occur when their sensitized crystals are irradiated at low temperatures. They attribute the absorption bands which appear to definite centres, which they assume to represent the earliest stages in the formation of photolytic silver (Stasiw 1951, 1952, Schöne 1953, Scholz 1954, 1956; see, however, Volke 1954). The absorption maximum at 7500 Å observed by Scholz (1954) when crystals of silver chloride containing silver sulphide are irradiated at liquid air temperature with light of wavelength 4050 Å is of particular interest. It was bleached with the appearance of green fluorescence when the crystals were irradiated with any wavelength in the range 6500 Å–7200 Å.

Clark and Mitchell (1956) found that crystals of silver chloride and silver bromide were more strongly sensitized, for the formation of internal photolytic silver by exposure at room temperature, by the incorporation of the corresponding cuprous halide than by the incorporation of silver sulphide.

10.3. *The Orientation of the Silver produced by Electron Bombardment and Photolysis with respect to the Silver Halide*

10.3.1. *General remarks.*

Dankov (1939) based a theory of the formation of the latent image and of chemical development upon the almost exact correspondence between the length of the face diagonal of the unit cell in silver with the edge of the cubic unit cell in silver bromide. He pointed out that this would allow silver to grow on silver bromide with an almost perfect match of the lattice periodicities at the interface when contact was made on (100) planes, those in one crystal being rotated through 45° with respect to those in the other. Dankov believed that this was responsible for the unique position held by silver bromide in photographic sensitivity. His speculations have stimulated many investigations of the possible relative orientations of silver with respect to silver chloride and silver bromide, in which :

(1) thin, oriented films of silver and silver halides have been deposited one on the other by vacuum evaporation,

(2) silver has been attacked by halogens to form thin films of the silver halide, and

(3) silver halides have been decomposed to produce silver by ultra-violet irradiation or electron bombardment.

The situation, which was at first confused, has now been partially clarified.

10.3.2. *Experimental observations.*

Kooy and Burgers (1948) first established by x-ray diffraction that the lattice of the silver produced by ultra-violet irradiation can be oriented with respect to the silver chloride lattice. Berry (1949) observed a parallel orientation between the silver bromide lattice and that of silver deposited upon a (111) surface by vacuum evaporation. The silver produced in the silver bromide crystal by ultra-violet irradiation was oriented with {110} planes in the silver rotated through 90° with respect to {110} planes in the silver halide. This was confirmed by Berry and Griffith (1950) by reflection electron-diffraction studies. They found that the silver, which was formed during ultra-violet irradiation near the surface of the crystal, grew with {110} planes in contact in this relationship. The {110} planes involved were mainly those which are inclined at 35° with respect to the (111) surface. At the surface, there was no evidence for the parallel orientation observed to result from vacuum deposition. Berry and Griffith found, however, that this was the only relationship between the silver and the silver halide within the crystals revealed by x-ray diffraction. Here, the symmetry directions in the silver lattice were always parallel to the corresponding directions in the silver bromide. Nye (1948 a, b) had shown that the glide planes in silver chloride always contained a ⟨110⟩ direction so that the edge dislocations could lie along ⟨110⟩ directions and have a Burgers vector parallel to ⟨110⟩. Berry and Griffith supposed that threads of silver would form along the edge dislocation lines in ⟨110⟩ directions, by the trapping of the electrons and the neutralization by silver ions of the trapped charges. They thus provided the first indirect evidence for the separation of silver along dislocation lines in crystals of silver bromide. Berry and Griffith (1953) next observed that the silver produced by reducing the surfaces of silver bromide crystals with very dilute developers was formed in the parallel orientation, whereas more rapid development resulted in random orientation and much twinning.

Trillat (1951, 1952 a, b, 1953) then studied the problem by electron-diffraction measurements in transmission, and found only random orientation of the silver produced by electron bombardment and development with respect to the silver halide. He has since found some parallel orientation (see Pashley 1952 d).

The orientation of the silver produced by bombardment with 30 to 60 kv electrons with respect to the silver halide matrix has been investigated by Pashley (1950, 1951, 1952 a, b, c, d), who used the reflection method and found that the silver is highly oriented in the parallel orientation in silver chloride and silver bromide, and randomly oriented in silver iodide. The silver produced on silver chloride by ultra-violet irradiation was also formed in the parallel orientation. There is always extensive twinning of the silver on (111) planes. In none of his work has the non-parallel orientation reported by Berry and Griffith (1950) been observed. Sonoike (1953, 1956) has also made observations on the orientation of photolytic silver in silver bromide. Besides the parallel orientation, he has observed the orientation found by Berry and Griffith (1950). He has, however, found other orientation relationships, and it is evident that the problem has not yet been fully resolved. The orientation originally proposed by Dankov (1939) has never been observed.

The only general conclusion which can be drawn from this experimental work is that the silver produced by photolysis with ultra-violet light and by electron bombardment in crystals of silver chloride and silver bromide is formed mostly in the parallel orientation. This is consistent with the formation of the internal silver nuclei by the separation of silver atoms along edge dislocations.

10.4. *The Theory of Photolysis*

The processes involved in the separation of visible specks of silver in pure and in sensitized crystals of silver chloride and silver bromide have already been discussed in §§ 4.5, 7.4.4 and 7.5. The mechanism for the separation of internal silver which has been proposed by Seitz (1951, 1952, 1956) relies upon the effective half-unit positive charge of silver ions occupying jogs along edge dislocation lines, for the trapping of conduction electrons. This mechanism has to be reconsidered when the partial homopolar character of the silver halides is taken into account, because it then appears unlikely that silver ions on jogs or at kink sites on the external surface will provide stable traps for electrons at room temperature. Furthermore, if this were the mechanism, the resulting effective half-unit negative charge would attract a positive hole rather than an interstitial silver ion, as has been pointed out in § 4.5, and recombination would follow. Seitz has also proposed that positive holes are trapped at jogs occupied by halide ions along edge dislocation lines, and he has envisaged the accumulation of halogen in the dislocations during photolysis. Visible specks of silver which do not recombine with halogen on storage cannot separate in the crystals unless positive holes are trapped at the external surfaces to form halogen molecules which escape as in the experiments of Luckey (1953, 1955).

According to Mitchell (1957 a, b), the same processes are involved in photolysis as in the formation of internal latent image specks. These have been considered in detail in §§ 7.4.4 and 7.5, and nothing more need be added here. The mechanism is essentially the same as that advanced originally by Gurney and Mott (1938), except that it is postulated that the adsorption of silver ions to the specks precedes the trapping of electrons (see § 7.5).

Mitchell postulates that the visible specks of silver are charged positively by the adsorption of silver ions when they are in thermal equilibrium with the crystal at room temperature, and that the compensating negative charge is provided by vacant silver ion lattice sites. This postulate leads to a satisfactory interpretation of the Becquerel effect, and to a partial resolution of the problems raised by Bodenstein (1940, 1941) and others (Noddack, Eckert and Meier 1952) in connection with the minimum quantum energy required to cause photochemical decomposition. After the formation of nuclei of silver, the photolytic decomposition of silver chloride and silver bromide can be continued by the absorption of quanta of lower energy than those which are required to initiate the decomposition. This is called the Becquerel effect, and it can be interpreted as follows. Both the unoccupied and occupied energy levels in the positively charged specks are lowered with respect to the conduction band and the full band of the crystal, compared with the positions of these levels in smaller, uncharged specks. In these circumstances, electrons may be transferred to the specks from occupied levels in the full band and from occupied surface levels in the forbidden band above the full band, by the absorption of quanta of smaller energy than are required for their transfer to the conduction band. The positive charge is then restored to the resulting uncharged specks by the adsorption of further silver ions and the creation of vacant silver ion lattice sites. The positive holes are trapped by halide *ions* at surface kink sites, and interstitial silver ions then diffuse to the neighbourhood of the specks to neutralize the space charge, as in § 7.4.4.

To explain why the minimum energy of the quanta which will just cause photochemical decomposition is considerably less than the value which would be estimated from the lattice energy determined by thermochemical methods, Bodenstein (1940, 1941) postulated that silver nuclei are always present in crystals of silver halides and that the quantum energy is reduced by the energy liberated when the silver atoms are adsorbed to these nuclei. The energy involved in a quantum process can, however, never be reduced by the liberation of energy in a secondary process which occurs at a later time. The heat of adsorption is liberated and the free energy of the system is reduced when the silver ion is adsorbed by the silver speck, before the absorption of the light quantum, with the creation of a vacant silver ion lattice site. This provides a partial solution of the problems raised by Bodenstein (1941), but further experimental work is needed to establish the longest wavelength of the light which will cause the formation of nuclei of silver in crystals initially free from nuclei. Large single crystals of silver chloride and silver bromide, which are as near mechanically perfect and as free from impurities as possible, have a vanishingly small sensitivity for the formation of silver nuclei when they are exposed to light with a wavelength near the long-wave limit of the lattice absorption band.

10.5. *The Principles underlying the Sensitization of Ionic Crystals for Photochemical Change*

The sensitization of ionic crystals for photochemical change has been discussed in many papers published during the past twenty years by Pohl and former members of his research school at Göttingen. With their experimental work on large single crystals, they have clearly defined the essential property of a sensitizing impurity. The impurity must trap the positive holes which are liberated with the electrons when energy is absorbed in the crystals. This means that the impurity must introduce occupied electronic energy levels into the forbidden band of the crystal, so that an impurity absorption appears beyond the long-wave limit of the lattice absorption band.

These properties are best exemplified by mixed crystals of potassium bromide and potassium hydride which have a well-defined impurity absorption band near the long wave edge of the potassium bromide absorption band. The sensitizing action of the hydride ions is clearly shown by the work of Martienssen (1951, 1952) on the formation of F-centres by exposure to x-rays, the significance of which has also been emphasized by Martienssen and Pohl (1952) and by Martienssen and Pick (1952, 1953, 1955). The quantum yield in the sensitized crystals in which the positive holes are trapped by the hydride ions is much greater than in the non-sensitized crystals where many positive holes and electrons recombine.

Sensitizing impurity centres must, however, also satisfy a further requirement which has not usually been sufficiently emphasized when the conclusions drawn from experimental work with large, single alkali halide crystals have been applied to the discussion of the problems of photographic sensitivity. To prevent regression, the group which results from the trapping of the positive hole by the sensitizing impurity centre must dissociate within the total excitation time of the electron (Mitchell 1954, 1956, 1957 a, b). The trapping of the positive hole is thus a necessary, but not a sufficient, condition for a high photochemical quantum efficiency. A further change must occur to prevent the combination of the conduction electron with the resulting positively charged centre.

In the case of potassium bromide crystals sensitized with potassium hydride, the centre produced by the trapping of the positive hole dissociates with the loss of a hydrogen atom which diffuses away. This removes the low-lying unoccupied energy level in the forbidden band and replaces it with an unoccupied level below the bottom of the conduction band associated with the vacant halide ion lattice site.

With the large crystals of silver chloride and silver bromide sensitized with sulphide, selenide and telluride ' anions ', which have been studied by Stasiw and his co-workers, the important sensitizing impurity centres in the crystals, in the thermally unstable state which results from quenching from temperatures near the melting point to room temperature, are probably association complexes between the divalent anions and interstitial silver ions. These centres can trap positive holes, leaving monovalent anions at the lattice sites, and then dissociate by the diffusion of interstitial silver ions away from the sites. The same state would result from the absorption of energy in the absorption band of the impurity anion. The absorption of energy in the crystals thus results in the liberation of conduction electrons and interstitial silver ions which then combine to form the groups of silver atoms which are responsible for the ' latent image ' absorption band.

Stasiw (1951, 1952, 1954, 1955) is of the opinion that these groups of silver atoms are distributed uniformly throughout the volumes of the sensitized crystals, and that association complexes between divalent anions and vacant halide ion lattice sites are involved in their formation. Our own experience with mixed crystals of silver halides and silver sulphide in the form of thin sheets (Clark and Mitchell 1956, Mitchell 1957 d) suggests that consideration should also be given to the possibility that the conduction electrons and interstitial silver ions combine, as in plastically deformed non-sensitized crystals, to form groups of silver atoms on internal surfaces provided by edge dislocation lines and other crystal imperfections.

§ 11. DEVELOPMENT

11.1. *The Surface and the Internal Latent Image*

The techniques introduced by Berg, Marriage and Stevens (1941 a, b, see also Stevens 1942) for distinguishing between the initiation of development by the surface and by the internal latent image are now well established. These, or similar methods, have been used in many investigations on the effects of chemical sensitization and the conditions of exposure on the distribution of the latent image, in recent years (Hautot 1950, Lowe, Jones and Roberts 1951, Hautot and Sauvenier 1953 b, c, Eggert and Okamoto 1956 a, b). Stevens (1951, 1953 a) has described improved methods for oxidizing the surface latent image before internal development, and James, Vanselow and Quirk (1953) have introduced more active surface developers containing metol and d-araboascorbic acid. Berg and Ford (1949) have worked out methods for differentiating between the surface and the internal latent image, using physical development. All these methods were used at Bristol for the experimental work on the sensitivity of large microcrystals and thin-sheet crystals of silver halides. As far as development is concerned, the same phenomena are observed whether reduction is initiated by a surface or by an internal latent image, so that no further reference will be made to the distribution of the latent image in this section.

11.2. *The Kinetics of Development*

11.2.1. *General remarks.*

The kinetics of development present three problems of great interest, towards an understanding of which much progress has been made during the past ten years :

(1) How does the developer differentiate between exposed and unexposed microcrystals ?

(2) What factors determine the induction period required for the development of the latent sub-image ?

(3) What factors determine the rate of development and the form in which the silver is produced after the end of the induction period ?

The present position with regard to the solution of these problems will be outlined in this section.

Diffusion effects are always of considerable importance in determining the kinetics of development of coated photographic emulsions and may lead to observations which do not lend themselves to immediate theoretical interpretation. It is therefore of interest to note that Fortmiller and James have investigated the kinetics of development of liquid photographic emulsions, in which the effects of diffusion are relatively unimportant, with p-phenylene diamine (1951), hydroquinone (1952) and metol (1953).

Reviews on the subject of developers and development have been published by James (1950, 1955), Abribat (1952), Levenson (1954), and Levenson and James (1954).

11.2.2. *Factors which influence the induction period.*

It is well known that the induction period for the development of the latent image is longer after a high than after a low intensity exposure. Many latent sub-image specks are formed at high intensities, whereas only a few latent image specks are formed at low intensities (Berg 1948). The induction period for the development of the latent sub-image may be reduced by post exposure and by chemical treatments. These methods are grouped under the general heading of latensification, and will be considered in the next section. The induction period is also influenced by the adsorption of many substances at the surfaces of the silver halide crystals. The particular case of development accelerators will be considered in § 11.2.3; the effect of other substances will be discussed here.

Gelatin, sensitizing dyes, desensitizing dyes, antifoggants and stabilizers, development accelerators, excess halide ions and the developing agents, may all be adsorbed at the surfaces of the silver halide crystals during development. The adsorption is in most cases reversible, and one agent may displace another under proper conditions and influence the kinetics of development. The composition of the adsorption layer is determined by the dynamic competition between these different ions and molecules which may be present in the environment of the grains. Gelatin is present throughout the preparation of the emulsion, and a layer of gelatin surrounds the silver halide grains in their final form. James (1940) has shown that a monolayer of adsorbed gelatin markedly decreases the rate of reduction of exposed and unexposed silver halide crystals, and similar observations were made by Evans, Hedges and Mitchell (1955). The gelatin may be displaced from the surfaces of the crystals by a number of strongly adsorbed cyanine sensitizing dyes (West, Carroll and Whitcomb 1952, 1953), but these dyes may themselves reduce the rate of chemical development. James and Vanselow (1951) have shown that allyl thiourea, which is strongly adsorbed, displaces all dyes from the surfaces of crystals of silver bromide. It should therefore also displace gelatin from the surfaces and can indeed increase the rate of initiation of development in certain acid developers. Thiosulphate ions also displace a number of merocyanin dyes from the surfaces of silver bromide crystals (James and Vanselow 1954). The displacement of an adsorbed merocyanine dye from the surfaces of particles of silver bromide by hydroquinone has also been observed by James and Vanselow (1952); they have thus established the adsorption of this developing agent to silver bromide which is of importance in connection with the triple-interface theory of development (see § 11.4.1). In general, the adsorption of an ion with the same charge will displace an adsorbed agent from the surface, whereas an ion of the opposite charge will promote adsorption. As many of the most widely used developing agents are active in the form of anions, the fundamental importance of the studies of co-operative and competitive adsorption with chromatographic columns filled with silver bromide particles, which have been undertaken by James and Vanselow (1951, 1952, 1954), for the discussion and interpretation of researches on the kinetics of development, will be appreciated.

11.2.3. *Latensification.*

The induction period for the development of a latent sub-image may be reduced by a number of preliminary chemical treatments. The density which is produced by development for a short period is thereby increased, so that there is an apparent gain in the speed of the photographic material. These treatments are described as latensification.

Latensification, by pre-bathing before development in solutions of sodium per-borate (Vanselow, Quirk and Leermakers 1948), and by exposure to ozone (Vanselow, Colten and Wiegand 1950), has been reported. The most effective method of latensification consists in bathing exposed emulsions, which have not previously been gold-sensitized, in solutions of complex salts of monovalent gold, before development (James 1948, James, Vanselow and Quirk 1948). This treatment reduces the induction period and increases the rate of subsequent development. Gold latensification eliminates high intensity reciprocity failure, greatly reduces the Herschel effect, and makes the latent sub-image more resistant to chemical attack by oxidizing agents.

Results obtained by James and Vanselow (1949) have shown that latensification is only effective in increasing the speed of photographic materials when weak developing solutions are employed, or when development with solutions of high activity is not carried to completion. The effects of latensification are clearly in evidence in the early stages of development but disappear on prolonged development. Latensification does not increase the number of grains which may be developed by energetic developers: it simply increases the rate at which at least some of the grains develop.

11.2.4. *Development accelerators.*

The induction period for the development of a latent sub-image may be substantially reduced and the contrast increased, by the addition of quaternary salts such as n-dodecyl pyridinium p-toluenesulphonate (Vanselow and James 1953) and non-ionic surface active agents such as polyoxyalkylene ethers (Blake 1950) to photographic developers. Vanselow and James (1953) have shown by their experiments with chromatographic columns that a number of quaternary salts, which act as development accelerators, displace adsorbed carbocyanine and merocyanine dyes from the surfaces of silver bromide particles. As these dyes will themselves displace gelatin from the particles, it may be concluded that the quaternary salts will displace gelatin. Direct evidence for this effect is difficult to obtain by experiments with chromatographic columns because gelatin seriously interferes with their operation. The displacement of gelatin from the surfaces of silver bromide crystals has, however, been demonstrated with a considerable number of development accelerators, by direct microscopic observation (Evans, Hedges and Mitchell 1955). There seems to be little doubt that surface-active agents which are effective accelerators of photographic development do displace gelatin and adsorbed sensitizing dyes from the surfaces of the crystals during development. This would decrease the induction period for the development of a latent sub-image, but this is certainly not their only action and may not be the most important one. James has shown that the cationic development accelerators promote the adsorption of anionic developing agents: his conclusions will be outlined in § 11.4.1.

These methods of latensification and development acceleration are incapable of increasing the sensitivities of modern high-speed negative emulsions which have been gold-sensitized and treated in other ways to ensure maximum efficiency in the development of the latent image.

11.3. *The Form of the Silver produced by Development*

Many active developers such as amidol (Hall and Schoen 1941) and p-aminophenol (Levenson and Tabor 1952) produce tangled bundles of filaments of silver when they reduce microcrystals of silver bromide. Other less active developers such as hydroquinone and metol–hydroquinone convert the microcrystals into masses of silver with a shape closely resembling that of the original halide (Evans and Mitchell 1955, Evans, Hedges and Mitchell 1955).

The study of the form of the silver produced by the reduction of crystals of silver halides with photographic developers lends itself to model experiments, in which large microcrystals and single crystals are treated with developing solutions and the progress of reduction after the induction period is followed with the microscope, either visually or photographically with a cine camera.

Some of the earliest experiments along these lines were made by Dankov and Kochetkov (1940) with thin sheets of silver bromide produced by crystallizing the molten material between glass plates. Development was initiated by fog specks on the surfaces of the crystals, and patches of silver with geometrical shapes were formed: these they believed to be single crystals of silver. They concluded that the observations supported the theory of development which had just previously been advanced by Dankov (1939, see also § 10.3.1).

The period since 1947 has been marked by a great increase in the number of model experiments on photographic development. These were initiated by Boissonnas (1949 a, b, 1950, 1951), who studied the reduction of the surfaces of polished slabs of silver bromide cut from large single crystals grown from the melt by the Kyropoulos method. He found that development began at fog specks on the surfaces of the crystals, and that dilute solutions of hydroquinone produced masses of silver in cavities which were bounded by {111} crystallographic planes. This work was continued at Neuchâtel by Vuille (1951, 1954 a, b), who found that masses of silver were formed in cavities bounded by both {111} and {100} planes, when very dilute developers were employed. Leidheiser and Cook (1954) have studied the reduction of highly polished single-crystal spheres of silver chloride, and observed characteristic differences in the rate of development on different crystallographic surfaces.

An extensive investigation of the influence of different variables on the form of the silver resulting from the reduction of thin sheets of silver bromide produced by the crystallization of the molten material between glass plates was then carried out by Keith and Mitchell (1953, see also Mitchell 1955, 1957 a, b). They studied the form of the silver produced by a wide range of commercial and experimental developers with the optical and the electron microscopes. By x-ray diffraction and by direct observation with the electron microscope, they showed that the specks of silver which are produced by the action of weak reducing agents on silver bromide crystals are not orientated single crystals of silver, as was assumed by Dankov and Kochetkov (1940) and by Boissonnas (1949 a, b), but compact aggregates of randomly oriented, exceedingly fine crystallites of silver. The effect on the form of the silver of additives, and of variations in physico-chemical conditions, was investigated, and the conditions in which loose, filamentary bundles and compact aggregates of fine crystallites appeared, were established. These experiments were supplemented by many other model experiments on the electrolytic reduction of crystals of silver bromide : these showed that the important site for reduction was the interface between the silver and the silver halide. The rate of development was drastically reduced by the addition of traces of many antifoggants which either formed very slightly soluble silver salts or were strongly adsorbed to metallic silver. From these experiments, it was concluded that the rate of development was determined by the availability of silver ions from kink sites on the surface of the silver halide crystal at its interface with the silver, and by the accessibility to silver ions of kink sites on the corresponding surfaces of the silver crystallites.

This work showed that the electrode mechanism proposed by Gurney and Mott (1938) and modified by Mott (1939, 1946, 1948) and by Berg (1943, 1948) to account for the formation of filamentary silver could play only a minor if any role in normal chemical development. No observations were made which required, for their interpretation, the assumption that interstitial silver ions or vacant silver ion lattice sites were involved in the development process. Keith and Mitchell (1953) concluded that silver ions were adsorbed to the silver nuclei at their interface with the silver halide or with the solution, that electrons were transferred to the silver nuclei from developer molecules, and that halide ions passed into solution, from the surface of the crystal or through a thin film of solution between the silver and the silver halide, to neutralize the space charges which would otherwise arise.

Jaenicke, Krüger and Hauffe (1951), Jaenicke and Schott (1955) and Jaenicke, Tischer and Gerischer (1955) have studied the behaviour of silver–silver bromide electrodes in developing solutions, and also concluded that electrode processes are involved in development.

11.4. *Theories of Development*

11.4.1. *The triple-interface theory.*

James (1943, 1944, 1946 a, b, 1950) has expressed the view that development is a catalysed reaction which takes place at the triple interface between the silver nuclei, the silver halide crystal and the solution. It involves the approach of the developer ions or molecules to the silver halide surface, and their adsorption to the silver halide to form complexes : these then break down into silver atoms and oxidation products of the developer, after their adsorption to the silver nucleus. According to this view, exposed and unexposed grains are differentiated by the occurrence of a catalysed reaction in the exposed grains, at the interface between the silver nuclei and the silver halide.

The kinetic studies which have been undertaken by James and his co-workers have shown that the rate of development is often dependent upon a fractional power of the concentration of the developing agent and this is most readily interpreted if the adsorption of the developing agent precedes reaction. Developing agents are seldom adsorbed to silver nuclei (James 1950), but they are adsorbed to the surfaces of silver halide particles (James and Vanselow 1952, see § 11.2.2). James has accumulated a large amount of evidence to support his theory and there is no doubt that adsorption plays an important role in chemical development.

The induction period for the development of a latent image is longer for developers with trivalent reducing anions than for those with monovalent anions, and it disappears when developers with neutral reducing anions are employed. According to James, the induction period arises from the fact that the reducing anions have difficulty in overcoming the potential barrier at the surfaces of the crystal due to the adsorption of halide and hydroxyl ions and the presence of an equivalent diffuse space charge of cations in the solution near the interface. The adsorption of some cationic dye molecules and of cationic development accelerators such as α-picolinium-β-phenylethyl bromide and n-dodecyl pyridinium p-toluenesulphonate reduces the induction period. Their adsorption at the surface of the crystal causes the diffuse space charge to collapse so that the reducing anions can approach the surface and be co-operatively adsorbed in accordance with the requirements of the triple interface mechanism. There are, however, also non-ionic surface active agents which have an influence on the kinetics of development (Blake 1950). The charge barrier theory can be applied to these substances only if an intra-molecular potential barrier is established through the adsorption of one end of a polarized bond to the silver halide surface while the other end projects into the solution. Cationic development accelerators tend to displace gelatin from silver halide surfaces whereas anionic surface active agents neither displace gelatin nor accelerate development. It may therefore be deduced that a non-ionic surface active agent which has a regular succession of polarized bonds along the chain and acts as a development accelerator should be adsorbed with the atoms with anionic character in contact with the silver halide and those with cationic character projecting into the solution. Gelatin molecules should then be displaced from the surface and an intra-molecular charge barrier established. Although this situation should facilitate the escape of halide ions from the surface of the crystal into the solution, it should equally well hinder the access of reducing anions to the surface : it is therefore difficult to see how these substances can act as development accelerators with anionic reducing agents if the charge barrier mechanism is the only important one responsible for the induction period. That more than one mechanism is involved is also shown by the decrease in the induction period for development with sodium hydroquinone monosulphonate which results from gold latensification (James 1950). If gold latensification merely converts latent image specks of silver into specks of gold, the decrease in the induction period can scarcely be interpreted simply in terms of the charge barrier hypothesis because the charge barrier should not be modified by the process.

11.4.2. *The Gurney–Mott theory of development.*

In their discussions of the electrode theory of development, Gurney and Mott (1938), Mott (1939, 1946, 1948), Berg (1943, 1948), Bagdasar'yan (1943, 1944, 1945) and Jaenicke and Schott (1955) have all followed the basic ideas of the Gurney–Mott theory of photolysis and of the formation of the latent image in assuming that the latent image speck is first charged negatively during development by the transference of electrons from reducing anions or neutral molecules and that the negatively charged speck then attracts silver ions from the crystal or from adjacent surface sites and discharges them with the formation of silver atoms at the interface between the silver nucleus and the silver halide. The balance of electrostatic charges in the system is maintained by the diffusion of vacant silver ion lattice sites away from the silver nucleus to surface sites, from which halide ions can escape into the solution. According to the theory, exposed and unexposed crystals are differentiated by the fact that the unoccupied energy levels in the uncharged silver nuclei lie below the conduction band of the silver halide crystal. During development, electrons are therefore transferred to these nuclei in exposed crystals, but not to the conduction band of the silver halide in the unexposed crystals.

There are many difficulties with this theory. The unoccupied electronic energy levels associated with latent sub-image specks consisting of pairs of silver or of gold atoms must lie above the bottom of the conduction band of the silver halide crystal (see § 7.2.4), so that the above mechanism for differentiation between exposed and unexposed grains during development cannot apply to grains bearing latent sub-image specks. The adsorption of molecules of the developer to the surface of the silver halide does not play any essential role and, although no specific discussion of the reasons for the induction period has ever been presented, it is clear that it would have to follow the lines of the James charge-barrier theory, except that the potential barrier at the surface of the crystal would prevent the developer anions from approaching the silver nuclei to charge them negatively, instead of hindering their approach to the silver halide surface and their adsorption there.

The postulate that lattice defects are involved in development is not supported by any observations made by Keith and Mitchell (1953) during their work on the mechanism of development. According to the Gurney–Mott theory of development either in its original form or as modified by Berg, the silver should be formed on the otherwise smooth surface of the silver halide, because only Frenkel defects are involved in the process. The experimental work shows clearly that the silver is formed in a cavity, in the surface of the crystal, produced by the escape of halide ions into the solution, along the interface between the silver and the silver halide.

11.4.3. *A revised electrode theory of development.*

The mechanisms which have been proposed by Mitchell for photochemical change in crystals of silver halides are distinguished from the Gurney–Mott mechanisms by the important feature that the adsorption of a silver ion to a latent image speck either accompanies or precedes the trapping of the electron (Mitchell 1956, 1957 a, b). The presence or adsorption of the silver ion lowers the unoccupied electronic energy levels associated with the speck, so that it becomes capable of trapping an electron. By applying these ideas to the development process, Mitchell has been able to arrive at a theory of the induction period which overcomes most of the difficulties already mentioned.

The latent sub-image specks consist of pairs of silver atoms adsorbed on the external and internal surfaces of the sensitivity centres. These specks are unable to trap electrons, either from the conduction band of the silver halide crystal during exposure, or from reducing molecules during development, unless they first adsorb silver ions from the crystal. They have, however, only a very small lifetime in the positively charged state at room temperature, because of the small heat of adsorption of silver ions to them and of the very low concentration of free silver ions available during development. If an electron is not transferred from a reducing ion or molecule near the speck during this lifetime, the group simply dissociates and reverts to the uncharged state. This explains why a considerable induction period may elapse before the occurrence of any change in the properties of the latent sub-image specks, as was established by Kornfeld and James (1943) in their studies of the Herschel effect (see § 9.2). The induction period is therefore influenced by any factor which modifies the concentration of silver ions or of reducing molecules at the interface between the latent image speck and the silver halide crystal. Many antifoggants which form slightly soluble silver salts are strongly adsorbed at sites adjacent to surface kink sites occupied by silver ions : they increase the induction period by decreasing the concentration of silver ions available for adsorption to latent sub-image specks. The adsorption of surface active agents (see § 11.2.4) may decrease the induction period by increasing : (i) the availability of silver ions from kink sites by displacing adsorbed sensitizing dyes and gelatin, (ii) the concentration of silver ions at the interface by promoting the diffusion of halide ions away from the interface, and (iii) the concentration of reducing molecules by removing the charge barrier and by the formation of co-operatively adsorbed layers. The transference of an electron to a positively charged latent sub-image speck produces an uncharged group of three silver atoms : this can adsorb a silver ion from the surface of the crystal as discussed in § 7.4.4, to form a stable positively charged group of four silver atoms in a tetrahedral arrangement. When such groups have been formed either during exposure, during latensification by post-exposure, or toward the end of the induction period during development, further reduction of the silver halide proceeds rapidly. Electrons are transferred to the groups from reducing ions or molecules, and the positive charges are restored by the adsorption of further silver ions from the crystal. The balance of electrostatic charges in the system is maintained by the passage of halide ions into the solution from the surface of the crystal at its interface with the silver speck.

The differentiation between exposed and unexposed crystals is explained as in the triple-interface theory by the presence or absence of silver nuclei because silver ions can combine with electrons in the earliest stages of development only after their adsorption to silver nuclei.

Evans, Hedges and Mitchell (1955) have shown that the latent image is formed with higher probability at the sensitivity centres, where there is a higher density of kink and edge sites, than elsewhere on the surfaces of the crystals. This is important because, as already mentioned, the initial rate of development is determined by the availability of silver ions for adsorption to silver nuclei, and the least firmly bound silver ions on the surface are those which occupy the kink sites. The availability of these silver ions may be controlled by the adsorption at the kink sites of gelatin, and of antifoggants, and this accounts for the influence of these substances on the induction period for the development of the latent sub-image, and on the rate of subsequent development.

The theory gives a direct explanation of the effect of gold sensitization in decreasing the induction period for the development of a latent sub-image. Burrow and Mitchell (1954) have pointed out that the heat of adsorption of a silver ion to a latent sub-image speck containing a gold atom is greater than the heat of adsorption to a speck consisting solely of silver atoms because of the greater polarizability of the gold ion which increases the van der Waals interaction. Mitchell (1957 a, b) therefore attributes the reduction in the induction period in gold-sensitized or gold-latensified emulsions to the increased lifetime in the positively charged state of latent sub-image specks containing gold atoms; this increases the probability for the transference of electrons from developer ions and molecules to them. The replacement of the silver atoms of latent sub-image specks by gold atoms may occur after exposure because no evidence that gold sensitization modifies either the primary or the secondary processes in the formation of the latent image has yet been presented.

The experimental evidence advanced by James to support the charge barrier theory of the induction period is not in conflict with the present theory. The probability for the transference of electrons to latent sub-image specks during their lifetime in the positively charged state must depend on the concentration of developer molecules or ions in the neighbourhood of the specks. This concentration may be increased by the adsorption of cationic development accelerators which displace the negatively charged gelatin molecules from the surface and increase the adsorption of reducing anions.

The visible masses of silver which are present during the later stages of development may be charged positively or negatively, according to the physicochemical conditions in the system in the quasi-steady state which is established. In the first case, the rate-determining step is the slow transference of electrons from developer molecules of low activity, and compact aggregates of extremely small crystallites of silver are produced, the shapes of which correspond closely to those of the original silver halide microcrystals. Rapid development is initiated at a few sites on the surfaces of the crystals at the end of the induction period, and masses of silver, which are usually bounded by circular arcs, spread outwards from the sites. Silver ions are adsorbed to the silver at its interface with the silver halide, and charge it positively. The positive charge is then neutralized by the transference of electrons from reducing ions or molecules, and halide ions diffuse along the interface into the solution.

The mass of silver is charged negatively when many developing solutions of high activity are employed, and the rate-determining process is then the transference of silver ions and their adsorption by the silver. In these conditions, the negative charge is concentrated at the extremities of the particles, and its neutralization there by silver ions from the surface of the crystal leads to the growth of tangled bundles of filaments of silver.

It should be noted that, in all cases of development, the silver is not in intimate contact with the silver halide but is separated from it by a thin film of solution. Halide ions diffuse along this film into the solution, and silver ions diffuse across it to be adsorbed to the silver masses. Whether the silver is charged positively or negatively with respect to the silver halide, the adhesion between the two is largely due to electrostatic attraction.

11.5. *Conclusion*

The essential difference between the theories of James and of Mitchell appears to lie in whether the silver ion is transferred to and adsorbed by the silver nucleus as a silver ion without the developer molecule, or whether the developer molecule is adsorbed to the silver bromide surface to form a complex with a silver ion which is then adsorbed as such to the silver nucleus. In the latter case, according to James, the developer molecule is split off in an oxidized form and a silver atom remains on the nucleus. This is equivalent to stating that silver ions as such are not adsorbed to the silver nucleus, and that electrons are not transferred to the nucleus from molecules of the developer.

It is clear that, in this complex subject, aspects of both theories are probably valid and that a final synthesis must await the results of more detailed kinetic studies than have yet been undertaken. So many factors influence the rate at which development proceeds, both in its initial and later stages, and the form in which the silver is produced by development, that it is difficult to establish that the essential variables have been isolated in any particular series of experiments.

§ 12. General Conclusions

The developments which have occurred during the past ten years in the different fields of crystal physics have had a profound impact upon the subject of photographic sensitivity. Although it was based upon a particularly simple model, the analysis of the structure and properties of surfaces in terms of Frenkel steps and kink sites which was made by Burton and Cabrera (1949 a, b) led to many conclusions which could be applied to various aspects of surface physics and chemistry. Their discussion of the difficulties inherent in the two-dimensional nucleation theory resulted in the introduction of the screw dislocation as an important factor in the growth of crystals (Frank 1949 a, b, 1952 a, b, Burton, Cabrera and Frank 1951). Frank's theory of crystal growth then received support from the observations of growth spirals on the surfaces of crystals which were made by Griffin (1950), Verma (1951), Forty (1954) and Dekeyser and Amelinckx (1955). There was also a steady accumulation of theoretical knowledge of the properties of stationary configurations of dislocations (Nabarro 1952).

The idea that dislocations were involved in photographic sensitivity was introduced for the first time by Mitchell (1951 a, 1952) on a purely hypothetical basis ; effects which might be attributed to their presence had not then been observed. Hedges and Mitchell (1953 a) next showed that configurations of edge dislocations in silver halide crystals could be rendered visible by the separation of silver and, almost immediately afterwards (1953 b), found that the internal latent image was formed along edge dislocation lines.

The experimental work which followed led to the conclusion that crystal imperfections, screw and edge dislocations in particular, have important functions in photographic sensitivity. Screw dislocations are introduced by the deformation of crystal nuclei ; they determine the growth of microcrystals during Ostwald ripening. Edge dislocations are introduced to relax internal strains arising from the growth of microcrystals upon imperfect nuclei, from the inclusion of foreign materials, and from the presence of abrupt concentration gradients of silver iodide in silver bromo-iodide crystals ; with other internal imperfections, they are responsible for the internal sensitivity of the crystals. Internal latent image specks and visible particles of silver are formed during exposure by the separation of silver atoms along the edge dislocation lines.

When an emulsion is washed before chemical sensitization, the surfaces of the microcrystals pass into solution at a higher rate at the dislocations than elsewhere, so that terraced depressions appear. The terraces are formed by incomplete planes of *ions* and the enhanced reactivity of their kinked edges, compared with that of the stable crystallographic planes which form their surfaces, is of fundamental importance for photographic sensitivity. The edges of the terraces are decorated with adsorbed molecules of silver sulphide during the early stages of sulphur sensitization and, from these edges, adsorbed monolayers of silver sulphide molecules spread out over the surfaces. The molecules of the sensitizing dyestuffs are also first adsorbed along the edges of the terraces and aggregated monolayers of dye molecules then form over the surfaces.

The kinked edge of any surface terrace, with the adsorbed molecules of silver sulphide and of sensitizing dyestuff, has all the properties which are essential for photographic sensitivity. The highest density of terraces occurs, however, at the depressions where the dislocations meet the surface of the crystal; these are the most reactive areas and the silver sulphide and dye molecules tend to be concentrated there. For this reason, it is convenient to describe the crystal imperfection or dislocation, in which the internal latent image is formed, with the associated reactive surface area, where the surface latent image appears, as a sensitivity centre. As a result of the introduction of the new ideas on the relationship between dislocations and the reactivity of crystal surfaces, we arrive at a more useful model, than has previously been available, for the discussion of the preparation and sensitization of the emulsion and the formation and development of the latent image.

There are two stages in the formation of the latent image. In the primary stage, conduction electrons and interstitial silver ions are made available for the formation of latent image specks by the absorption of energy, either within the crystals, or by silver sulphide or dye molecules, adsorbed at their surfaces. The silver sulphide molecules either trap positive holes or lose electrons to the conduction band; the resulting positively charged molecules then dissociate into AgS molecules and interstitial silver ions. In the secondary stage, the interstitial silver ions and the conduction electrons combine, through a series of intermediate steps, to form latent image specks at sites adjacent to silver *ions* along the edges of surface terraces or along edge dislocation lines. The unstable latent pre-image specks, with a lifetime between 0·2 and 20 seconds, are silver atoms adsorbed at such sites. The latent sub-image specks, with a lifetime of days, are pairs of silver atoms, similarly adsorbed; they are transformed into uncharged latent image specks by combination with a third interstitial silver ion and conduction electron. The group of three silver atoms then adsorbs a silver ion from the surface of the crystal and forms a positively charged tetrahedrally disposed group of four silver atoms with a vacant silver ion lattice site to maintain electrical neutrality. The latent image consists of a distribution of these and larger positively charged specks on the external and internal surfaces of the crystals.

The review prepared ten years ago by Berg (1948) was based upon the Gurney–Mott theory of the formation of the latent image. It may therefore be of interest to conclude this review with a brief summary of the changes which have taken place during the intervening period. The latent image was formed in two stages according to the Gurney–Mott theory. In the primary stage, electrons and positive holes were liberated by the absorption of energy within the crystals; the electrons were then supposed to be trapped by uncharged sensitivity specks of silver sulphide or of silver and the positive holes to be trapped at other surface sites to form halogen which escaped. In the secondary stage, the negatively charged silver sulphide or silver specks attracted interstitial silver ions which combined with the electrons on the specks to form silver atoms. There are three main difficulties with the Gurney–Mott theory: (1) silver sulphide molecules and small groups of silver atoms are unlikely to trap conduction electrons because their electron affinity cannot exceed the large electron affinity (about 3·5 ev) of the silver halide crystal; (2) even if they did, a mobile positive hole would reach the negatively charged speck or group of atoms before an interstitial silver ion so that no silver atoms could separate; and (3) the neutral groups of silver atoms forming the latent image should themselves trap positive holes and electrons successively and thus merely serve as recombination centres for the electronic charge carriers.

Two stages are also involved in the theory of the formation of the latent image for which the author has been responsible, but the emphasis is placed on the trapping of positive holes during the primary stage and not on that of electrons; the adsorbed molecules of silver sulphide trap the positive holes to form positively charged molecules which dissociate into AgS molecules and interstitial silver ions. In the secondary stage, the interstitial silver ions and conduction electrons combine to form latent image specks, and this process is not directly influenced by the presence of silver sulphide molecules and other products of chemical sensitization.

The fundamental differences between the two theories arise from the positions which are assigned in them to the unoccupied (electron trapping) and the occupied (positive hole trapping) levels of the silver sulphide molecules, latent image specks, and visible silver particles, relative to the bottom of the conduction band and to the top of the full band.

Gurney and Mott placed the unoccupied (electron trapping) levels of the uncharged silver sulphide and silver specks near the middle of the forbidden band; this implies that a small group of molecules of silver sulphide or of atoms of silver has a very large electron affinity (about 4·5 ev). With the unoccupied levels in this position, the electrons can, however, be trapped during exposure to charge the groups negatively as is required by the theory. Gurney and Mott indicated in their energy level diagrams and in their interpretation of the Herschel effect that the silver specks had metallic properties with the Fermi level near the middle of the forbidden band but the trapping of positive holes by the higher occupied levels in the uncharged silver specks or in the silver sulphide sensitivity specks was never discussed. With the unoccupied energy levels in both silver sulphide sensitivity specks and latent image specks near the middle of the forbidden band, difficulties also arise in explaining the differentiation between unexposed and exposed crystals during development.

Mitchell (1956, 1957 a, b) has placed the unoccupied levels of silver sulphide molecules and small uncharged groups of silver atoms above the bottom of the conduction band corresponding to an electron affinity of about 1·5 ev; they can therefore neither trap conduction electrons nor receive electrons from reducing molecules during development. The occupied levels of adsorbed silver sulphide molecules and latent sub-image specks are placed about the middle of the forbidden band corresponding to the extension of the optical absorption to longer wavelengths which accompanies sulphur sensitization and to the wavelength at which the maximum Herschel effect occurs; both adsorbed molecules of silver sulphide and latent sub-image specks therefore trap positive holes and then dissociate with the loss of an interstitial silver ion.

In the new theory, the conduction electrons are not trapped by surface silver *ions* or by latent pre- or sub-image specks unless the corresponding unoccupied energy levels are first lowered by the presence of an interstitial silver ion at the trapping site; with this mechanism, a localized negative charge, which can attract positive holes, never appears. The unoccupied energy levels are raised again by the trapping of the electron and latent sub-image specks can consequently not accept electrons directly from reducing molecules during the early stages of development. This requires the adsorption of a silver ion to the latent sub-image speck to lower the unoccupied energy levels while a reducing molecule which can transfer an electron is in the immediate neighbourhood; the low probability for the occurrence of this situation is responsible for the induction period.

An uncharged latent image speck adsorbs a silver ion from the surface of the crystal to become a positively charged speck with a negatively charged vacant silver ion lattice site in the neighbourhood to maintain electrical neutrality. The positive potential arising from the adsorption of the silver ion lowers the unoccupied and the occupied levels of the speck below the bottom of the conduction band and the top of the full band, respectively. The positively charged specks do not therefore trap positive holes and are thus protected from regression; they trap conduction electrons during exposure and they accept electrons from reducing molecules during development. The transference of an electron raises the energy levels but they are then lowered again by the adsorption of a further silver ion from the surface of the crystal and the creation of a vacant silver ion lattice site. The new conceptions which have been introduced thus remove all the difficulties encountered in the application of the Gurney–Mott theory to the formation of the latent image and of visible particles of silver during exposure and to the explanation of the differentiation between unexposed and exposed crystals during development.

The Gurney–Mott theory was particularly fruitful in stimulating research in the field of latent image effects, and the fundamental principle that the motion of both electrons and silver ions is involved in the formation of the latent image is now firmly established. It is appropriate to refer to this as the Gurney–Mott principle because they were the first workers in the field to recognize its significance for photographic sensitivity. The principle is, however, consistent with many alternative mechanisms and it now seems clear that those originally proposed by Gurney and Mott for the formation of latent image and visible specks of silver require revision.

The new theory which has been presented in this paper extends beyond the formation of the latent image and the initiation of development and embraces most of the physical and chemical aspects of photographic sensitivity. It is based upon experimental observations and leads to a series of working hypotheses which should prove useful in the planning of future fundamental and technological researches. The subject is one of fascinating complexity which still presents many unsolved problems to attract both the physicist and the chemist.

REFERENCES

Reports of Photographic Conferences

Liège, 1948, Hautot, A., 1948, *Sci. Industr. Photogr.* (2), **19**, 281, 321, 361.
Zürich, 1949, Eggert, J., and Ammann, H., 1950, *Journées internationales de science photographique* (Paris : Revue d'Optique).
Bristol, 1950, Mitchell, J. W., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths†).
Paris, 1951, Morand, M., and Vassy, A., 1953, *Colloque sur la sensibilité des cristaux et des émulsions photographiques* (Paris : Revue d'Optique).
Hakone, 1953, Fujisawa, S., 1956, *Photographic Sensitivity* (Tokyo : Maruzen).
London, 1953, Schultze, R. S., 1955, *Science and Applications of Photography* (London : Royal Photographic Society).
Cologne, 1956, Frieser, H., 1957, *Wissenschaftliche Photographie* (Darmstadt : Helwich).

Useful Reference Works

COTTRELL, A. H., 1953, *Dislocations and Plastic Flow in Crystals* (Oxford : Clarendon Press).
DEKEYSER, W., and AMELINCKX, S., 1955, *Les dislocations et la croissance des cristaux* (Paris : Masson).
FRIEDEL, J., 1956, *Les dislocations* (Paris : Gauthier-Villars).
GARNER, W. E., 1955, *Chemistry of the Solid State* (London : Butterworths).
KITTEL, C., 1956, *Introduction to Solid State Physics*, 2nd Edn (New York : John Wiley).
MEES, C. E. K., 1954, *The Theory of the Photographic Process*, 2nd Edn (New York : Macmillan).
VASSY, A., 1953, *Fondements théoriques de la photographie* (Paris : Revue d'Optique).

Original Papers

ABRIBAT, M., 1952, *Photogr. J.* B, **92**, 25.
ALLCOCK, G. R., 1956, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **5**, 412.
ALLEMAND, C., and ROSSEL, J., 1954 a, *Helv. Phys. Acta*, **27**, 212 ; 1954 b, *Ibid.*, **27**, 519.
ALLEN, J. R., 1953, *Canad. J. Phys.*, **31**, 218.
AMMANN-BRASS, H., 1953, *Sci. Industr. Photogr.* (2), **24**, 399.
AMSLER, H., 1953, *Z. Elektrochem.*, **57**, 801.
ANASTASEVICH, V. S., and FRENKEL, J., 1941, *J. Exp. Theor. Phys.*, **11**, 127.
ARENS, H., 1948, *Z. wiss. Photogr.*, **43**, 120 ; 1954 a, *Veröff. wiss. Photo-Lab. Agfa*, **8**, 8 ; 1954 b, *Ibid.*, **8**, 17.
ARENS, H., EGGERT, J., and KLEINSCHROD, F. G., 1943, *Z. wiss. Photogr.*, **42**, 33.
BAGDASAR'YAN, KH. S., 1943, *J. Phys. Chem., Moscow*, **17**, 336 ; 1944, *Acta Phys.-Chim. U.R.S.S.*, **19**, 421 ; 1945, *Ibid.*, **20**, 441.

---

† To save space Butterworth's Scientific Publications is referred to throughout these references as Butterworths.

Baker, E. A., 1956, *J. Photogr. Sci.*, **4**, 101.

Ballard, A. E., Stevens, G. W. W., and Zuehlke, C. W., 1952, *Photogr. Soc. Amer. J.* B, **18**, 27.

Belliot, H., 1951, *Sci. Indust. Photogr.* (2), **22**, 81 ; 1952, *Ibid.* (2), **23**, 52.

Berg, W. F., 1939, *Trans. Faraday Soc.*, **35**, 445 ; 1940, *Proc. Roy. Soc.* A, **174**, 559 ; 1943, *Trans. Faraday Soc.*, **39**, 115 ; 1947, *Photogr. J.* B, **87**, 112 ; 1948, *Rep. Progr. Phys.* (London : Physical Society), **11**, 248 ; 1952, *Schweiz. Photordsch.*, **17**, 38, 57, 77 ; 1953, *J. Photogr. Sci.*, **1**, 2.

Berg, W. F., and Burton, P. C., 1948, *Photogr. J.* B, **88**, 84.

Berg, W. F., and Ford, D. G., 1949, *Photogr. J.* B, **89**, 31.

Berg, W. F., Marriage, A., and Stevens, G. W. W., 1941 a, *J. Opt. Soc. Amer.*, **31**, 385 ; 1941 b, *Photogr. J.*, **81**, 413.

Berry, C. R., 1949, *Acta Cryst., Camb.*, **2**, 393 ; 1951, *Phys. Rev.*, **82**, 422 ; 1955, *Ibid.*, **97**, 676.

Berry, C. R., and Griffith, R. L., 1950, *Acta Cryst., Camb.*, **3**, 219 ; 1953, *Ibid.*, **6**, 94.

Berry, C. R., and Marino, S. J., 1955, *Photogr. Sci. Technique*, **2**, 149.

Biltz, M., 1949, *J. Opt. Soc. Amer.*, **39**, 994 ; 1951, *Z. Naturforsch.*, **6a**, 336 ; 1952, *J. Opt. Soc. Amer.*, **42**, 1952.

Birr, E. J., 1952 a, *Z. wiss. Photogr.*, **47**, 2 ; 1952 b, *Ibid.*, **47**, 72 ; 1953, *Ibid.*, **48**, 103 ; 1954 a, *Ibid.*, **49**, 1 ; 1954 b, *Ibid.*, **49**, 261 ; 1955 a, *Ibid.*, **50**, 107 ; 1955 b, *Ibid.*, **50**, 124.

Blake, R. K., 1950, U.S. Patent 2,533,990.

Bodenstein, M., 1940, *Naturwissenschaften*, **28**, 145 ; 1941, *Abh. Preuss. Akad. Wiss.*, No. 19, 1.

Boissonnas, C. G., 1949 a, *Sci. Industr. Photogr.* (2), **20**, 361 ; 1949 b, *Experientia*, **5**, 282 ; 1950, *C.R. Acad. Sci., Paris*, **230**, 1278 ; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 36.

Bourne, B. P., and Loening, E. E., 1953, *J. Photogr. Sci.*, **1**, 159.

Bradley, D. E., 1954, *Brit. J. Appl. Phys.*, **5**, 96.

Brooker, L. G. S., 1954, *The Theory of the Photographic Process* (New York : Macmillan), p. 371 ; 1955, *XIVth International Congress of Pure and Applied Chemistry, Zürich, Experientia Suppl.*, p. 229.

Brooker, L. G. S., White, F. L., Heseltine, D. W., Keyes, G. H., Dent, S. G., and Van Lare, E. J., 1953, *J. Photogr. Sci.*, **1**, 173.

Brooker, L. G. S., White, F. L., Sprague, R. H., Dent, S. G., and Van Zandt, G., 1947, *Chem. Rev.*, **41**, 325.

Brown, F. C., 1953, *Phys. Rev.*, **92**, 502 ; 1955, *Ibid.*, **97**, 355.

Burrow, J. H., and Mitchell, J. W., 1954, *Phil. Mag.*, **45**, 208.

Burton, P. C., 1946, *Photogr. J.* B, **86**, 62 ; 1948 a, *Ibid.*, **88**, 13 ; 1948 b, *Ibid.*, **88**, 123.

Burton, P. C., and Berg, W. F., 1946, *Photogr. J.* B, **86**, 2.

Burton, W. K., and Cabrera, N., 1949 a, *Disc. Faraday Soc.*, **5**, 33 ; 1949 b, *Ibid.*, **5**, 40.

Burton, W. K., Cabrera, N., and Frank, F. C., 1951, *Phil. Trans. Roy. Soc.* A, **243**, 299.

Capstaff, J. G., and Bullock, E. R., 1920, *J. Franklin Inst.*, **190**, 871.

Carroll, B. H., and Hubbard, D., 1931, *Bur. Stand. J. Res., Wash.*, **7**, 219 ; 1933, *Ibid.*, **11**, 743 ; 1934, *Ibid.*, **12**, 329.

Carroll, B. H., and West, W., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 162.

Chateau, H., and Pouradier, J., 1954, *Sci. Industr. Photogr.* (2), **25**, 3.

Chibisov, K. V., 1956, *Z. wiss. Photogr.*, **51**, 59.

Chibisov, K. V., Titov, A. A., and Mikhailova, A. A., 1951, *Doklady Akad. Nauk, S.S.S.R.*, **78**, 319 ; 1952, *Ibid.*, **84**, 547.

Clark, P. V. McD., and Mitchell, J. W., 1956, *J. Photogr. Sci.*, **4**, 1.

Damschroder, R. E., 1949, U.S. Patent 2,597,856.

Dankov, P. D., 1939, *Doklady Akad. Nauk, S.S.S.R.*, **24**, 773.

Dankov, P. D., and Kochetkov, A. A., 1940, *Doklady Akad. Nauk, S.S.S.R.*, **26**, 785.

Davies, K. N., and Holliday, A. K., 1952 a, *Trans. Faraday Soc.*, **48**, 1061 ; 1952 b, *Ibid.*, **48**, 1066.

Debot, R., 1941, *Bull. Soc. Roy. Sci. Liège*, p. 675 ; 1942 a, *Ibid.*, p. 126 ; 1942 b, *Mem. Soc. Roy. Sci. Liège* (4), **6**, 91.

Demers, P., 1950, *Canad. J. Res.* A, **28**, 628 ; 1952, *Sci. Industr. Photogr.* (2), **23**, 1 ; 1954, *Canad. J. Phys.*, **32**, 538.

Dickinson, H. O., 1948, *Photogr. J.* B, **88**, 97 ; *Ibid.*, **90**, 142 ; 1954, *J. Photogr. Sci.*, **2**, 50.

Ebert, I., and Teltow, J., 1955, *Ann. Phys., Lpz.* (6), **15**, 268.

Eggert, J., 1926, *Z. Elektrochem.*, **32**, 491.

Eggert, J., and Biltz, M., 1938, *Trans. Faraday Soc.*, **34**, 892 ; 1939, *Veröffentl. wiss. Zentral-Lab. Photogr. Abt. Agfa*, **6**, 23 ; 1940, *Z. wiss. Photogr.*, **39**, 140.

Eggert, J., and Kleinschrod, F. G., 1939, *Veröffentl. wiss. Zentral-Lab. Photogr. Abt. Agfa*, **6**, 37 ; 1940 a, *Z. wiss. Photogr.*, **39**, 155 ; 1940 b, *Ibid.*, **39**, 165.

Eggert, J., Meidinger, W., and Arens, H., 1948, *Helv. Chim. Acta*, **31**, 1163.

Eggert, J., and Okamoto, Y., 1956 a, *Z. Elektrochem.*, **60**, 71 ; 1956 b, *Ibid.*, **60**, 500.

Eggert, J., and Rahts, W., 1928, *Handbuch der Physik* (Berlin : Springer), **19**, 606.

Evans, C. H., and Hirschlaff, E., 1939, *J. Opt. Soc. Amer.*, **29**, 164.

Evans, T., Hedges, J. M., and Mitchell, J. W., 1955, *J. Photogr. Sci.*, **3**, 73.

Evans, T., and Mitchell, J. W., 1955, *Defects in Crystalline Solids* (London : Physical Society), p. 409.

Fajans, K., and Frankenburger, W., 1923, *Z. phys. Chem.*, **105**, 255.

Fajans, K., and Karagunis, G., 1929, *Z. phys. Chem.* B, **5**, 385.

Fajans, K., and Steiner, W., 1927, *Z. phys. Chem.*, **125**, 307.

Falla, L., 1952, *Sci. Industr. Photogr.* (2), **23**, 48 ; 1953 a, *Ibid.* (2), **23** A, 46 ; 1953 b, *Ibid.* (2), **23** A, 128 ; 1954, *Ibid.* (2), **25**, 348 ; 1955 a, *Ibid.* (2), **26**, 3 ; 1955 b, *Ibid.* (2), **26**, 81.

Farnell, G. C., 1952, *Phil. Mag.*, **43**, 289.

Farnell, G. C., and Burton, P. C., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 61.

Farnell, G. C., Burton, P. C., and Hallama, R., 1950, *Phil. Mag.*, **41**, 157, 545.

Fleischmann, R., 1930, *Ann. Phys., Lpz.* (5), **5**, 73.

Fortmiller, L., and James, T. H., 1951, *Photogr. Soc. Amer. J.*, **17** B, 102 ; 1952, *Ibid.*, **18** B, 76 ; 1953, *Ibid.*, **19** B, 109.

Fortmiller, L., James, T. H., Quirk, R. F., and Vanselow, W., 1950, *J. Opt. Soc. Amer.*, **40**, 487.

Forty, A. J., 1954, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **3**, 1.

Franck, J., and Livingstone, R., 1949, *Rev. Mod. Phys.*, **21**, 505.

Franck, J., and Teller, E., 1938, *J. Chem. Phys.*, **6**, 861.

Frank, F. C., 1949 a, b, *Disc. Faraday Soc.*, **5**, 48, 67 ; 1952 a, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **1**, 91 ; 1952 b, *Z. Elektrochem.*, **56**, 429.

Frenkel, J., 1931 a, b, *Phys. Rev.*, **37**, 17, 1276.

Friauf, R. J., 1954, *J. Chem. Phys.*, **22**, 1329.

Frieser, H., and Klein, E., 1955, *Mitteil. Forschungslab. Agfa Leverkusen-München*, **1**, 1.

Fujisawa, S., and Mizuki, E., 1951, *Bull. Soc. Sci. Photogr. Japan*, **14**, 9.

Fujisawa, S., Mizuki, E., and Kubotera, K., 1957, *Wissenschaftliche Photographie* (Darmstadt : Helwich).

Gay, P., Hirsch, P. B., Thorp, J. S., and Kellar, J. N., 1951, *Proc. Phys. Soc.* B, **64**, 374.

Gilleo, M. A., 1953 a, *Phys. Rev.*, **90**, 364 ; 1953 b, *Ibid.*, **91**, 534.

Goddard, P. E., and Urbach, F., 1952, *J. Chem. Phys.*, **20**, 1975.

Grenishin, S. G., and Minkov, I. M., 1954, *Doklady Akad. Nauk, S.S.S.R.*, **96**, 459.

Griffin, L. J., 1950, *Phil. Mag.*, **41**, 196.

Grimley, T. B., 1950 a, *Thesis*, University of Bristol ; 1950 b, *Proc. Roy. Soc.* A, **201**, 40 ; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 46.

Grimley, T. B., and Mott, N. F., 1947, *Disc. Faraday Soc.*, **1**, 3.

Gurney, R. W., and Mott, N. F., 1938, *Proc. Roy. Soc.* A, **164**, 151.

Hall, C. E., and Schoen, A. L., 1941, *J. Opt. Soc. Amer.*, **31**, 281.

Hamilton, J. F., Hamm, F. A., and Brady, L. E., 1956, *J. Appl. Phys.*, **27**, 874.

HAMM, F. A., and COMER, J. J., 1953, *J. Appl. Phys.*, **24**, 1495.

HAUTOT, A., 1945, *Bull. Soc. Roy. Sci. Liège*, p. 200 ; 1948, *Sci. Industr. Photogr.* (2), **19**, 281 ; 1950, *Ibid.* (2), **21**, 241 ; 1951 a, *Bull. Soc. Roy. Sci. Liège*, **20**, 249 ; 1951 b, *Ibid.*, **20**, 400.

HAUTOT, A., and DEBOT, R., 1948, *Photogr. J.* B, **88**, 139.

HAUTOT, A., and FALLA, L., 1942, *Bull. Soc. Roy. Sci. Liège*, **11**, 298.

HAUTOT, A., and SAUVENIER, H., 1949, *Sci. Industr. Photogr.* (2), **20**, 286 ; 1951 a, *Ibid.* (2), **22**, 201 ; 1951 b, *Z. Naturf.*, **6a**, 370 ; 1953 a, *Sci. Industr. Photogr.* (2), **24**, 232 ; 1953 b, *Ibid.* (2), **24**, 257 ; 1953 c, *Ibid.* (2), **24**, 429 ; 1955, *Ibid.* (2), **26**, 201 ; 1957 a, *Ibid.* (2), **28**, 1 ; 1957 b, *Ibid.* (2), **28**, 57.

HAYNES, J. R., and SHOCKLEY, W., 1948, *Report on Conference on Strength of Solids* (London : Physical Society), p. 151 ; 1951, *Phys. Rev.*, **82**, 935.

HECHT, K., 1932, *Z. Phys.*, **77**, 235.

HEDGES, J. M., and MITCHELL, J. W., 1953 a, *Phil. Mag.*, **44**, 223 ; 1953 b, *Ibid.*, **44**, 357.

VAN HEERDEN, P. J., 1945, *Thesis*, University of Utrecht ; 1950, *Physica*, **16**, 505.

VAN HEERDEN, P. J., and MILATZ, J. M. W., 1950, *Physica*, **16**, 517.

HERZ, R. H., and GROUNSELL, G. V., 1957, *Wissenschaftliche Photographie* (Darmstadt : Helwich).

HICKMAN, K. C. D., 1927, *Photogr. J.*, **67**, 34.

HILSCH, R., and POHL, R. W., 1930, *Z. Phys.*, **64**, 606 ; 1932, *Ibid.*, **77**, 421.

HIRSCH, P. B., and KELLAR, J. N., 1951, *Proc. Phys. Soc.* B, **64**, 369.

HOERLIN, H., and MUELLER, F. W. H., 1950, *J. Opt. Soc. Amer.*, **40**, 246.

HOLLIDAY, A. K., and DAVIES, K. N., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 56.

HUSE, E., and MEULENDYKE, C. E., 1926, *Photogr. J.*, **66**, 306.

JAENICKE, W., KRUGER, A., and HAUFFE, K., 1951, *Z. phys. Chem.*, **197**, 161.

JAENICKE, W., and SCHOTT, C., 1955, *Z. Elektrochem.*, **59**, 956.

JAENICKE, W., TISCHER, R. P., and GERISCHER, H., 1955, *Z. Elektrochem.*, **59**, 448.

JAMES, T. H., 1940, *J. Amer. Chem. Soc.*, **62**, 536, 1649, 1654, 3411 ; 1943, *J. Chem. Phys.*, **11**, 338 ; 1944, *Ibid.*, **12**, 453 ; 1946 a, *Ibid.*, **14**, 536 ; 1946 b, *J. Chem. Educ.*, **23**, 595 ; 1948, *J. Coll. Sci.*, **3**, 447 ; 1950, *Advances in Catalysis* (New York : Academic Press), **2**, 105 ; 1955, *Science and Applications of Photography* (London : Royal Photographic Society), p. 155.

JAMES, T. H., and VANSELOW, W., 1949, *Photogr. Soc. Amer. J.*, **15**, 688 ; 1951, *J. Amer. Chem. Soc.*, **73**, 5617 ; 1952, *Ibid.*, **74**, 2374 ; 1953 a, *J. Photogr. Sci.*, **1**, 133 ; 1953 b, *J. Phys. Chem.*, **57**, 725 ; 1954, *Ibid.*, **58**, 894.

JAMES, T. H., VANSELOW, W., and QUIRK, R. F., 1948, *Photogr. Soc. Amer. J.*, **14**, 349 ; 1953, *Ibid.*, **19** B, 170.

JELLEY, E. E., 1936, *Nature, Lond.*, **138**, 1009 ; 1937, *Ibid.*, **139**, 378, 631.

JENNY, L., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 259.

JOHNSTON, W. G., 1955, *Phys. Rev.*, **98**, 1777.

JOST, W., 1933, *J. Chem. Phys.*, **1**, 466 ; 1938, *Trans. Faraday Soc.*, **34**, 860.

JOST, W., and MENNENÖH, S., 1950, *Z. phys. Chem.*, **196**, 188.

JOST, W., MENNENÖH, S., and MÜLLER, F. H., 1949, *Z. Naturf.*, **4a**, 227.

JOST, W., and NEHLEP, G., 1936 a, *Z. phys. Chem.* B, **32**, 1 ; 1936 b, *Ibid.*, **34**, 350.

JUNG, L., STASIW, O., and TELTOW, J., 1951, *Z. phys. Chem.*, **198**, 186.

KAISER, W., 1952, *Z. Phys.*, **132**, 497.

KANZAKI, H., 1951, *Phys. Rev.*, **81**, 884 ; 1956, *Photographic Sensitivity* (Tokyo : Maruzen), p. 15.

KATZ, E., 1949, *J. Chem. Phys.*, **17**, 1132 ; 1950, *Ibid.*, **18**, 499.

KEITH, H. D., and MITCHELL, J. W., 1951, *Phil. Mag.*, **42**, 1331 ; 1953, *Ibid.*, **44**, 877.

KLEIN, E., 1954, *Photogr. Korr.*, **90**, 52.

KOCH, E., and WAGNER, C., 1937, *Z. phys. Chem.* B, **38**, 295.

KOOY, J. N., and BURGERS, W. G., 1948, *Rec. Trav. Chim. Pays-Bas*, **67**, 21.

KORNFELD, G., 1949, *J. Opt. Soc. Amer.*, **39**, 490.

Kornfeld, G., and James, T. H., 1943, *J. Opt. Soc. Amer.*, **33**, 615.

Koslowsky, R., 1951, *Z. wiss. Photogr.*, **46**, 65 ; 1953, *Photogr. Korr.*, **89**, 205.

Kossel, W., 1927, *Nachr. Ges. Wiss. Göttingen*, 135.

Kreveld, A. van, and Jurriens, H. J., 1937, *Physica*, **4**, 285, 297.

Krummenerl, T., 1952, *Sci. Industr. Photogr.* (2), **23** A, 171 ; 1955, *Science and Applications of Photography* (London : Royal Photographic Society), p. 86.

Kuhn, H., 1948, *Helv. Chim. Acta*, **31**, 1141.

Kurnick, S. W., 1952, *J. Chem. Phys.*, **20**, 218.

Larsen, E. T., Mueller, F. W. H., and Hoerlin, H., 1953, *J. Phys. Chem.*, **57**, 802.

Leermakers, J. A., Carroll, B. H., and Staud, C. J., 1937, *J. Chem. Phys.*, **5**, 878.

Lehfeldt, W., 1935, *Nachr. Ges. Wiss. Göttingen*, **1**, 171.

Leidheiser, H., and Cook, F. H., 1954, *J. Amer. Chem. Soc.*, **76**, 617.

Lescynski, W., 1926, *Z. wiss. Photogr.*, **24**, 261, 275.

Levenson, G. I. P., 1954, *J. Photogr. Sci.*, **2**, 72.

Levenson, G. I. P., and James, T. H., 1954, *J. Photogr. Sci.*, **2**, 169.

Levenson, G. I. P., and Tabor, J. H., 1952, *Sci. Industr. Photogr.* (2), **23** A, 295.

Loening, E. E., 1951 a, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 126 ; 1951 b, *Ibid.*, p. 149 ; 1952, *Photogr. J.* B, **92**, 126 ; 1953, *Sci. Industr. Photogr.* (2), **23** A, 199.

Loening, E. E., Sutherns, E. A., and Bourne, B. P., 1954, *J. Photogr. Sci.*, **2**, 160.

Lowe, W. G., and Jones, J. E., 1948, U.S. Patent 2,604,397.

Lowe, W. G., Jones, J. E., and Roberts, H. E., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 112.

Lottermoser, A., 1907, *Z. phys. Chem.*, **60**, 451 ; 1908, *Ibid.*, **62**, 284, 359.

Luckey, G. W., 1953, *J. Phys. Chem.*, **57**, 791 ; 1954, *Photogr. Sci. Technique*, **1**, 115 ; 1955, *J. Chem. Phys.*, **23**, 882.

Luckey, G. W., and West, W., 1956, *J. Chem. Phys.*, **24**, 879.

Maerker, R. E., 1954, *J. Opt. Soc. Amer.*, **44**, 625.

Martienssen, W., 1951, *Naturwissenschaften*, **38**, 482 ; 1952, *Z. Phys.*, **131**, 488.

Martienssen, W., and Pick, H., 1952, *Nachr. Akad. Wiss. Göttingen* A, **11**, 117 ; 1953, *Z. Phys.*, **135**, 309 ; 1955, *Science and Applications of Photography* (London : Royal Photographic Society), p. 110.

Martienssen, W., and Pohl, R. W., 1952, *Z. Phys.*, **133**, 153.

Mees, C. E. K., 1955, *Science and Applications of Photography* (London : Royal Photographic Society), p. 7.

Meidinger, W., 1940, *Phys. Z.*, **41**, 277 ; 1944, *Fortschr. Photogr.*, **3**, 1 ; 1949 a, b, *Z. wiss. Photogr.*, **44**, 1, 137.

Meier, H., 1954, *Z. Elektrochem.*, **58**, 859.

Meiklyar, P. V., 1946, *Doklady Akad. Nauk, S.S.S.R.*, **54**, 705.

Meiklyar, P. V., and Putseiko, E. K., 1950, *Doklady Akad. Nauk, S.S.S.R.*, **73**, 63.

Mennenöh, S., 1950, *Z. Elektrochem.*, **54**, 433.

Mitchell, J. W., 1948, *Sci. Industr. Photogr.* (2), **19**, 361 ; 1949 a, *Phil. Mag.*, **40**, 249 ; 1949 b, *Ibid.*, **40**, 667 ; 1951 a, *Bull. Soc. Roy. Sci. Liège*, **20**, 300 ; 1951 b, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 242 ; 1952, *Sci. Industr. Photogr.* (2), **23**, 457 ; 1953 a, *Ibid.* (2), **23** A, 213 ; 1953 b, *J. Photogr. Sci.*, **1**, 110 ; 1954, *Z. Phys.*, **138**, 381 ; 1955 a, *Science and Applications of Photography* (London : Royal Photographical Society), pp. 68, 69, 73 ; 1955 b, *Chemistry of the Solid State* (London : Butterworths), p. 311 ; 1956, *Z. Elektrochem.*, **60**, 557 ; 1957 a, *J. Photogr. Sci.*, **5** ; 1957 b, *Die photographische Empfindlichkeit* (Darmstadt : Verlag Helwich) ; 1957 c, *Dislocations and Mechanical Properties of Crystals* (New York : John Wiley) ; 1957 d, *Wissenschaftliche Photographie* (Darmstadt : Helwich).

Mizuki, E., and Fujisawa, S., 1956, *Photographic Sensitivity* (Tokyo : Maruzen), p. 97.

Moffitt, W. E., 1950, *Proc. Phys. Soc.* A, **63**, 700.

Moser, F., and Urbach, F., 1956, *Phys. Rev.*, **102**, 1519.

Mott, N. F., 1939, *Rep. Progr. Phys.*, **6**, 186 (London : Physical Society) ; 1941, *Photogr. J.*, **81**, 62 ; 1946, *J. Phys. Radium*, **7**, 249 ; 1948, *Photogr. J.* B, **88**, 119.

Mott, N. F., and Gurney, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford: Clarendon Press).

Mueller, F. W. H., 1949, *J. Opt. Soc. Amer.*, **39**, 494; 1955, *Science and Applications of Photography* (London: Royal Photographic Society), p. 13.

Nabarro, F. R. N., 1952, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **1**, 269.

Nail, N. R., Moser, F., and Urbach, F., 1956, *J. Opt. Soc. Amer.*, **46**, 218.

Nelson, R. C., 1956 a, b, c, *J. Opt. Soc. Amer.*, **46**, 10, 13, 1016.

Noddack, W., and Meier, H., 1953, *Z. Elektrochem.*, **57**, 691.

Noddack, W., Eckert, G., and Meier, H., 1952, *Z. Elektrochem.*, **56**, 735.

Nye, J. F., 1948 a, *Nature, Lond.*, **161**, 367; 1948 b, *Ibid.*, **162**, 299.

Oh-yama, Y., 1947 a, *J. Soc. Sci. Japan*, **10**, 115; 1947 b, *Ibid.*, **11**, 2; 1956, *Photographic Sensitivity* (Tokyo: Maruzen), p. 133.

Okamoto, Y., 1956, *Nachr. Akad. Wiss. Göttingen*, No. 14, 275.

Pashley, D. W., 1950, *Acta Cryst., Camb.*, **3**, 163; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London: Butterworths), p. 39; 1952 a, *Acta Cryst., Camb.*, **5**, 850; 1952 b, *Proc. Phys. Soc.* A, **65**, 33; 1952 c, *Sci. Industr. Photogr.* (2), **23** A, 17; 1952 d, *Phil. Mag.*, **43**, 1208.

Pick, H., 1951, *Naturwissenschaften*, **38**, 323.

Pick, H., and Pohl, R. W., 1951, *Z. Naturf.*, **6a**, 360.

Pitts, E., 1954, *J. Chem. Phys.*, **22**, 56.

Pouradier, J., and Chateau, H., 1954, *Sci. Industr. Photogr.* (2), **25**, 188.

Pouradier, J., and Roman, J., 1952, *Sci. Industr. Photogr.* (2), **23**, 4.

Randall, J. T., 1939, *Trans. Faraday Soc.*, **35**, 2.

Raynaud, H., and Pouradier, J., 1955 a, *J. Chim. Phys.*, **52**, 133; 1955 b, *Sci. Industr. Photogr.* (2), **26**, 206.

Rexer, E., 1931, *Z. Phys.*, **70**, 159; 1932 a, *Ibid.*, **75**, 777; 1932 b, *Ibid.*, **76**, 735.

Russell, G., 1956, *J. Photogr. Sci.*, **4**, 94.

Sauvenier, H., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London: Butterworths), p. 215; 1954, *Sci. Industr. Photogr.* (2), **25**, 271.

Scheibe, G., 1936, *Z. angew. Chem.*, **49**, 563; 1937 a, *Ibid.*, **50**, 51, 212; 1937 b, *Naturwissenschaften*, **25**, 75, 474, 795; 1938, *Koll. Z.*, **82**, 1; 1948, *Z. Elektrochem.*, **52**, 283; 1952, *Ibid.*, **56**, 723.

Scheibe, G., and Kandler, L., 1938, *Naturwissenschaften*, **26**, 412.

Scholz, A., 1954, *Z. Phys.*, **137**, 207; 1956, *Ann. Phys., Lpz.* (6), **19**, 175.

Schöne, E., 1953, *Z. Phys.*, **136**, 52.

Schöne, E., Stasiw, O., and Teltow, J., 1951, *Z. phys. Chem.*, **197**, 145.

Seifert, G., and Stasiw, O., 1955, *Z. Phys.*, **140**, 97.

Seifert, G., Stasiw, O., and Volke, C., 1954, *Naturwissenschaften*, **41**, 58.

Seitz, F., 1950 a, *Phys. Rev.*, **79**, 890; 1950 b, *Ibid.*, **79**, 1002; 1950 c, *Ibid.*, **79**, 1003; 1950 d, *Ibid.*, **80**, 239; 1951, *Rev. Mod. Phys.*, **23**, 328; 1952, *Imperfections in Nearly Perfect Crystals* (New York: John Wiley), p. 3; 1956, *Photographic Sensitivity* (Tokyo: Maruzen), pp. 5, 136.

Shapiro, I., and Kolthoff, I. M., 1947, *J. Chem. Phys.*, **15**, 41.

Sheppard, S. E., 1925 a, *Photogr. J.*, **65**, 380; 1925 b, *Colloid Symposium Monograph*, **3**, 76; 1925 c, British Patent 235,211; 1926, *Brit. J. Photogr.*, **73**, 33; 1928, *Photogr. J.*, **68**, 397.

Sheppard, S. E., and Graham, C. L., 1940, *J. Franklin Inst.*, **230**, 619.

Sheppard, S. E., Lambert, R. G., and Walker, R. D., 1939, *J. Chem. Phys.*, **7**, 265, 426; 1941, *Ibid.*, **9**, 96.

Sheppard, S. E., Trivelli, A. P. H., and Loveland, R. P., 1925, *J. Franklin Inst.*, **200**, 51.

Sheppard, S. E., Trivelli, A. P. H., and Wightman, E. P., 1927, *Photogr. J.*, **67**, 281.

Sheppard, S. E., Vanselow, W., and Hall, V. C., 1929, *J. Phys. Chem.*, **33**, 1403.

Sheppard, S. E., Vanselow, W., and Happ, G. P., 1940, *J. Phys. Chem.*, **44**, 411.

Simpson, J. H., 1949, *Proc. Roy. Soc.* A, **197**, 269; 1955, *Ibid.*, **231**, 308.

Simpson, W. T., 1947, *J. Chem. Phys.*, **15**, 414.

Slade, R. E., and Toy, F. C., 1920, *Proc. Roy. Soc.* A, **97**, 181.

SMEKAL, A., 1932, *Bericht über der VIII Internationalen Kongress, Dresden*, 1931 (Leipzig : Barth), p. 34.

SMITH, L. P., 1951, *Semi-Conducting Materials*, Ed. H. K. Henisch (London : Butterworths), p. 114.

SPENCE, J., and CARROLL, B. H., 1948, *J. Phys. Chem.*, **52**, 1090.

SONOIKE, S., 1953, *J. Appl. Phys. Japan*, **22**, 19 ; 1956, *Photographic Sensitivity* (Tokyo : Maruzen), p. 59.

STASIW, O., 1949 a, *Journées internationales de science photographique* (Paris : Revue d'Optique), p. 48 ; 1949 b, *Ann. Phys., Lpz.* (6), **5**, 151 ; 1950, *Z. Phys.*, **127**, 522 ; 1951, *Ibid.*, **130**, 39 ; 1952 a, *Ibid.*, **134**, 106 ; 1952 b, *Z. Elektrochem.*, **56**, 749 ; 1953, *Sci. Industr. Photogr.* (2), **23** A, 12 ; 1954, *Z. Phys.*, **138**, 246 ; 1955, *Halbleiterprobleme* (Braunschweig : Vieweg), **2**, 184.

STASIW, O., and TELTOW, J., 1941, *Ann. Phys., Lpz.* (5), **40**, 181 ; 1944, *Nachr. Akad. Wiss. Göttingen*, 155 ; 1951, *Z. Naturf.*, **6a**, 363.

STAUDE, H., 1956, *Z. Elektrochem.*, **60**, 563.

STAUFFER, R. E., and SMITH, W. F., 1948, U.S. Patent 2,598,079.

STEIGMANN, 1924, *Chem. Zeit.*, **46**, 234 ; 1928, *Koll. Zeitschr.*, **46**, 57 ; 1930, *Ibid.*, **53**, 318 ; 1951, *Ibid.*, **122**, 171 ; 1952 a, *Sci. Industr. Photogr.* (2), **23**, 256 ; 1952 b, *Ibid.* (2), **23**, 354 ; 1953 a, *Ibid.* (2), **24**, 179 ; 1953 b, *Ibid.* (2), **24**, 217 ; 1953 c, *Ibid.* (2), **24**, 386 ; 1954, *Ibid.* (2), **25**, 190 ; 1955, *Ibid.* (2), **26**, 289 ; 1956, *Ibid.* (2), **27**, 223.

STEVENS, G. W. W., 1942, *Photogr. J.*, **82**, 42 ; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 227 ; 1953 a, *J. Photogr. Sci.*, **1**, 122 ; 1953 b, *Sci. Industr. Photogr.* (2), **23** A, 194.

STRANSKI, I. N., 1928, *Z. phys. Chem.*, **136**, 259.

SUPTITZ, P., 1957, *Wissenschaftliche Photographie* (Darmstadt : Helwich).

SUTHERNS, E. A., 1956, *J. Photogr. Sci.*, **4**, 83.

SUTHERNS, E. A., and LOENING, E. E., 1956 a, *J. Photogr. Sci.*, **4**, 148 ; 1956 b, *Ibid.*, **4**, 154.

SUZUKI, S., 1956, *Photographic Sensitivity* (Tokyo : Maruzen), p. 39.

TAFT, E., and APKER, L., 1957, *Phys. Rev.*, in the press.

TELTOW, J., 1949 a, *Ann. Phys., Lpz.* (6), **5**, 63 ; 1949 b, *Ibid.* (6), **5**, 71 ; 1950 a, *Z. phys. Chem.*, **195**, 197 ; 1950 b, *Ibid.*, **195**, 213 ; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London : Butterworths), p. 33 ; 1952 a, *Sci. Industr. Photogr.* (2), **23**, 302 ; 1952 b, *Z. Elektrochem.*, **56**, 767 ; 1956, *Halbleiterprobleme* (Braunschweig : Vieweg), **3**, 26.

TOY, F. C., and HARRISON, G. B., 1930, *Proc. Roy. Soc.* A, **127**, 613, 629.

TRILLAT, J. J., 1951, *C.R. Acad. Sci., Paris*, **233**, 1188 ; 1952 a, *Acta Cryst., Camb.*, **5**, 471 ; 1952 b, *Sci. Industr. Photogr.* (2), **23**, 129 ; 1953, *C.R. Acad. Sci., Paris*, **236**, 60.

TRIVELLI, A. P. H., 1909, *Arch. neerland. Sci. exakt. natur.* (2), **14**, 339 ; 1923, *Rec. Trav. Chim. Pays-Bas*, **42**, 714 ; 1930, *J. Franklin Inst.*, **209**, 373.

TRIVELLI, A. P. H., and SMITH, W. F., 1939, *Photogr. J.*, **79**, 330 ; 1940, *Ibid.*, **80**, 285.

TUBANDT, C., 1932, *Wien–Härms Handbuch d. Experimentalphysik* (Leipzig : Akad. Verlags.), **12**, 384.

TUTIHASI, S., 1957, *Phys. Rev.*, **105**, 882.

UMANO, S., 1956, *Photographic Sensitivity* (Tokyo : Maruzen), p. 129.

URBACH, F., 1935, *IXth International Photographic Conference* (Paris : Revue d'Optique), p. 432 ; 1953, *Phys. Rev.*, **92**, 1324.

URBACH, F., and URBACH, A., 1950, *Acta Phys. Austriaca*, **3**, 422.

URBACH, F., and WOLINSKI, A., 1938, *Wien. Akad. Ber.*, **147**, 31.

VANSELOW, W., COLTON, E. K., and WIEGAND, E. J., 1950, *Photogr. Soc. Amer. J.*, **16** B, 97.

VANSELOW, W., and JAMES, T. H., 1950, *Photogr. Soc. Amer. J.*, **16**, 21 ; 1953, *Ibid.*, **19** B, 36.

VANSELOW, W., QUIRK, R. F., and JAMES, T. H., 1953, *Photogr. Soc. Amer. J.*, **19** B, 66.

VANSELOW, W., QUIRK, R. F., and LEERMAKERS, J. A., 1949, *Photogr. Soc. Amer. J.*, **14**, 675.

VANSELOW, W., and SHEPPARD, S. E., 1929, *J. Phys. Chem.*, **33**, 331.

VASSY, A., and VASSY, E., 1948, *C.R. Acad. Sci., Paris*, **226**, 1183.

VERMA, A. R., 1951, *Phil. Mag.*, **42**, 1005.

VOLKE, C., 1954, *Z. Phys.*, **138**, 623 ; 1956, *Ann. Phys., Lpz.* (6), **19**, 203.

VUILLE, R., 1951, *Sci. Industr. Photogr.* (2), **22**, 219; 1954 a, *Helv. Chim. Acta*, **37**, 2264; 1954 b, *Ibid.*, **37**, 2271.

WAGNER, C., and BEYER, J., 1936, *Z. phys. Chem.*, B **32**, 113.

WALLER, C., COLLINS, R. B., and DODD, E. C., 1943, U.S. Patent No. 2,399,038.

WEBB, J. H., 1942, *J. Opt. Soc. Amer.*, **32**, 299; 1950 a, *Ibid.*, **40**, 3; 1950 b, *Ibid.*, **40**, 197; 1955, *J. Appl. Phys.*, **26**, 1309.

WEBB, J. H., and BERG, W. F., 1954, *The Theory of the Photographic Process* (New York: Macmillan), p. 204.

WEBB, J. H., and EVANS, C. H., 1938 a, *J. Opt. Soc. Amer.*, **28**, 249; 1938 b, *Ibid.*, **28**, 431; 1940, *Ibid.*, **30**, 445.

WEST, W., 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London: Butterworths), p. 99.

WEST, W., and CARROLL, B. H., 1947, *J. Chem. Phys.*, **15**, 529; 1951, *Ibid.*, **19**, 417; 1953, *J. Phys. Chem.*, **57**, 797; 1954, *The Theory of the Photographic Process* (New York: Macmillan), p. 430.

WEST, W., CARROLL, B. H., and WHITCOMB, D. H., 1952, *J. Phys. Chem.*, **56**, 1054; 1953, *J. Photogr. Sci.*, **1**, 145.

WEST, W., and SAUNDERS, V. I., 1957, *Wissenschaftliche Photographie* (Darmstadt: Helwich).

WOOD, H. W., 1952, *Sci. Industr. Photogr.* (2), **23**, 209; 1953, *J. Photogr. Sci.*, **1**, 163; 1954, *Ibid.*, **2**, 154; 1955 a, *Ibid.*, **3**, 144; 1955 b, *Ibid.*, **3**, 169.

YAMADA, K., and OKA, S., 1956, *Naturwissenschaften*, **43**, 175.

YAMAKAWA, K. A., 1951, *Phys. Rev.*, **82**, 522.

YAMANAKA, C., and SUITA, T., 1955, *Technol. Rep. Osaka Univ.*, **5**, 47.

YUTZY, H. C., and LEERMAKERS, J. A., 1949, U.S. Patent No. 2,597,915.

ZIMEN, K. E., 1946, *Arkiv. f. Kemi, mineralogi och Geologi* A, **23**, 1; 1951, *Fundamental Mechanisms of Photographic Sensitivity* (London: Butterworths), p. 53.

THEORY AND PROPERTIES OF SOLID ARGON

By E. R. DOBBS and G. O. JONES
Department of Physics, Queen Mary College, London

CONTENTS

PAGE
§ 1. Introduction 517
§ 2. Theory of crystal lattices 518
  2.1. Quantum-mechanical basis 518
  2.2. Interatomic potentials 520
  2.3. Equilibrium and stability of ideal crystals 523
    2.3.1. Mechanical stability of cubic and hexagonal lattices 523
    2.3.2. Absolute stability of close-packed lattices 526
  2.4. Dynamics of the cubic close-packed lattice 528
  2.5. Thermodynamics of ideal crystals 532
    2.5.1. General equation of state 532
    2.5.2. Law of corresponding states 535
    2.5.3. Zero-point effects 536
    2.5.4. Equations of state using model potentials 538
  2.6. Anharmonicity in ideal crystals 541
    2.6.1. Equation of state 541
    2.6.2. Thermal conduction 543
    2.6.3. Melting 545
§ 3. Properties of solid argon 546
  3.1. Crystallization and strength 547
  3.2. Zero-point properties 548
  3.3. Thermodynamic parameters 550
    3.3.1. Specific heat 550
    3.3.2. Density and expansivity 552
    3.3.3. Compressibility and elastic constants 554
    3.3.4. Grüneisen's constant 555
  3.4. Non-equilibrium properties 555
  3.5. Melting curve and sublimation curve 557
  3.6. Solid solution with other inert gas atoms 558
§ 4. Conclusion 559
Acknowledgments 560
Appendices : Tabulated properties of the inert gas crystals 560
References 562

Abstract. In the last few years measurements have been made of the thermodynamic properties of solid argon which provide ideal material for the verification of the theory of crystal lattices. Starting from the basic problem of the crystal lattice, a review is given here of the dynamics and thermodynamics of the inert gas crystals. Satisfactory agreement is obtained in many cases between theory and experiment when the calculations are based on an anharmonic Einstein theory with a (12,6) interatomic potential. It is shown that further progress requires, in particular, the extension of experiments to lower temperatures and of theoretical calculations of the repulsion of argon atoms to a higher approximation.

§ 1. INTRODUCTION

THE classic work of Born and his collaborators on the theory of the sodium chloride crystal was of unique importance in physics because it attempted, for the first time, a real atomic theory of the solid state. In particular it was shown that the property of cohesion could be related quantitatively to the electrostatic attraction between sodium and chlorine atoms, assumed to be completely ionized. The extension of Born's work to metals, as attempted, for example, by Grüneisen, depended on the use of interatomic potential functions known to have no direct physical significance in such solids. The same criticism applies to the calculation of the frequency spectra of metallic crystals, using the methods first applied by Blackman to the case of ionic crystals. On the other hand, the behaviour of pairs of inert gas atoms can be closely represented by interatomic potentials.

While the 'inert gas' crystals have thus been the subject of many theoretical investigations, these have had to rely on a knowledge of the properties of the substances in the gaseous state. In particular, the variation with temperature of the second virial coefficient has been used to provide values of the necessary potential constants. A better approach would have been to make use of lattice constants obtained directly, but until quite recently the measurements made of the properties of solidified helium, argon, etc., have been of insufficient accuracy.

From many points of view these solids are ideal subjects for investigations on the detailed dynamical behaviour of crystal lattices. They crystallize in the simple close-packed arrangements to be expected of atoms with spherical symmetry. Their interatomic forces consist mainly of the ordinary van der Waals attraction, which can be well represented by a simple radial potential, is short-range and is, therefore, essentially a two-body force. Further, we do not have to consider interactions between electrons and the lattice, nor, in general, the contributions of electrons to the physical properties. In comparison with the classical example of sodium chloride and other ionic substances the inert gas crystals have the additional advantages, for our purpose, that they are monatomic and that there are no complications due to the presence of strong electrostatic fields. Although, of course, they are essentially low-temperature solids and can only be studied at temperatures below the boiling-point of liquid air, this is not such a great disadvantage as might be thought. Many refined experimental techniques, such as those employing ultrasonic vibrations or X-rays, can be applied easily in low-temperature experiments, and the main difficulty is only that of preparing suitable specimens, since conventional handling techniques cannot be used. (For example, in ultrasonic experiments employing quartz crystals, the solid cannot be stuck to the crystal in the ordinary way, but can be deposited directly from the vapour upon the quartz surface.) If we are interested in the range of temperature immediately below the melting-point, then it is probably easier to carry out accurate measurements with the inert gas crystals than with any other materials.

Argon has been chosen for most of the experimental work because it can be obtained cheaply in a very pure state and because, as compared with helium and neon, it exists in the solid state over a practically convenient temperature range. From the point of view of establishing comparison between experiment and theory it is also an advantage that argon, again unlike helium and neon, has a zero-point energy which is only a small fraction of its total energy. In this report we shall give an account of the present state of the theory of the properties of solid argon, making use of the accurate experimental data recently obtained. We shall examine various theoretical treatments and discuss the use of various interatomic potentials. As a background to this discussion we consider first the basic ideas of crystal lattice theory.

§ 2. THEORY OF CRYSTAL LATTICES

The microscopic theory of crystal lattices is the solution of a many-body problem in which a very large, but finite, number of interacting particles are in motion about a set of ordered points. Classically, this set of points is the crystal lattice and each particle is at rest at a lattice point at the absolute zero. The early developments of the quantum theory of solids used different approximations to the complex vibrational motions of the particles and so are ' model ' rather than ' lattice ' theories. The first model due to Einstein (1907), in which all the particle vibrations have the same frequency ν_E, was quickly followed by the Debye (1912) model, in which the vibrations are those of an elastic continuum up to a certain limiting frequency ν_D. In the later model of Grüneisen (1926), the frequencies ν_i are not specified, but they all vary with the volume V in the same manner, so that the Grüneisen constant, $\gamma = -d\ln\nu_i/d\ln V$, depends on the volume only. Using any one of these three models, the general thermodynamic behaviour of solids can be studied without solving the dynamical, many-body problem of the crystal lattice.

2.1. *Quantum-mechanical Basis*

The fundamental problem of the crystal lattice is the specification of the motions of all the n nuclei and all the n_e electrons that together form the lattice. In quantum theory, these motions are derived from the Schrödinger equation for their wave functions

$$H\Psi = E\Psi, \qquad \ldots\ldots(1)$$

where H is the total Hamiltonian, E is the total energy and Ψ is the total wave function for the system. Now the Hamiltonian H is the sum of the kinetic energies of the nuclei T_n and electrons T_e and the total electrostatic energy V. If the n nuclei are of mass $m_k (k = 1, 2, \ldots, n)$ with position coordinates x_k and momenta p_k, and the n_e electrons are similarly specified by (m_{ej}, x_{ej}, p_{ej}), then $H = T_n + T_e + V$, that is,

$$H(x,p;\,x_e,p_e) = \tfrac{1}{2}\sum_k \frac{p_k^2}{m_k} + \tfrac{1}{2}\sum_j \frac{p_{ej}^2}{m_{ej}} + V(x, x_e). \qquad \ldots\ldots(2)$$

In all real crystals, the kinetic energies of the nuclei are very small compared with those of the electrons, since the masses of the nuclei are from 10^3 to 10^5 times the electronic mass. The motions of the electrons can therefore be found by considering the approximation (Born and Oppenheimer 1927) in which the nuclei are fixed, so that equation (1) becomes, for the electronic motion

$$(T_e + V)\phi = E^0\phi \qquad \ldots\ldots(3)$$

where E^0 denotes the energy of interaction of the system with fixed nuclei and ϕ is the electronic wave function. Both the eigenvalues and eigenfunctions will depend on the nuclear coordinates x_0 as parameters and the latter on the $3n_e$ electronic coordinates x_e. Hence they can be written

$$E^0 = \Phi_i(x_0); \quad \phi = \phi_i(x_0, x_e) \qquad \ldots\ldots(4)$$

where i is the electronic quantum number. The lowest eigenvalue (the ground state) of the total electrical (electronic plus nuclear electrostatic) energy is $\Phi_0(x_0)$.

For some crystals, all the electrons remain in this ground state under wide variations of temperature and pressure. It is then possible for any configuration (x, x_e) to use this lowest eigenvalue $\Phi_0(x)$ and the corresponding eigenfunction $\phi_0(x, x_e)$ to find the nuclear wave function $\chi(u)$, where u are the *small* displacements $(x-x_0)$ of the nuclei from their initial, fixed positions. That is, equation (1) becomes for the nuclear motion

$$(T_n + \Phi_0)\,\chi(u) = E\chi(u) \qquad \ldots\ldots(5)$$

where E, the total energy of the system, stays constant during the motion and so is independent of u, whereas $\Phi_0(x)$, the electrical energy, may be written as a Taylor series in u:

$$\Phi_0(x) = \Phi_0(x_0+u) = \Phi_0^{(0)}(x_0) + \Phi_0^{(1)}(u) + \Phi_0^{(2)}(u^2) + \Phi_0^{(3)}(u^3) + \ldots \qquad \ldots\ldots(6)$$

In this way, Φ_0 is used as the potential function for the nuclei to find their wave function $\chi(u)$ and the total wave function of the system becomes

$$\Psi(x, x_e) = \chi(u)\,\phi_0(x, x_e). \qquad \ldots\ldots(7)$$

This approximation for the total wave function is known as the *adiabatic approximation*, since the factor $\phi_0(x, x_e)$ implies that the electrons are moving as though the nuclei were fixed in their instantaneous positions. Transitions of electrons from the ground state to excited states are excluded, the nuclear motion merely deforming the ground state. It can be shown (see, for example, Born and Huang 1954) that this is a valid approximation for non-metallic lattices, where the energy difference ΔE between the first excited state and the ground state, for a particular nuclear configuration, satisfies the condition

$$\Delta E \gg h/2\pi\tau \qquad \ldots\ldots(8)$$

where h is Planck's constant and τ is the time interval in which the nuclear motion produces an appreciable change in $\phi_0(x, x_e)$. The approximation fails in the case of metallic lattices where the higher energy states overlap.

Non-metallic lattices may be molecular (He, Ne, A, ...; H_2, D_2; CH_4, etc.), or ionic (NaCl, AgBr, etc.), in which the molecules or stable ions are themselves saturated units. In other non-metallic solids (sulphur, diamond, etc.) there are no such units, but the electronic configuration is nevertheless saturated. It is evident that the use of a potential function, $\Phi_0(x)$ in (5), is appropriate only in the case of non-metallic lattices and the detailed study of nuclear motions in solids is thus best undertaken with non-metals.

The procedure is to solve the wave equation (5) by regarding the nuclear motion as causing small perturbations u from the equilibrium lattice x_0. In equilibrium, the intermolecular force must vanish at each lattice point and so

$$\Phi_0^{(1)}(u) = \sum_k \left(\frac{\partial\Phi_0(x)}{\partial x_k}\right)_{x_0} u_k = 0. \qquad \ldots\ldots(9)$$

The potential function $\Phi_0(x)$ thus consists of a static potential, $\Phi_0^{(0)}(x_0)$, independent of u, together with a series of terms beginning with $\Phi_0^{(2)}(u^2)$. When the nuclear motion is found using just this quadratic function of the nuclear displacements as the effective potential function, then the wave equation (5) becomes

$$(T_n + \Phi_0^{(2)})\,\chi(u) = \epsilon\chi(u). \qquad \ldots\ldots(10)$$

It has as solutions a set of simple harmonic oscillations with eigenvalues $\epsilon_j (j = 1, \ldots, 3n)$ given by

$$\epsilon_j = (n_j + \tfrac{1}{2})\,\hbar\omega_j \qquad \ldots\ldots(11)$$

where $\omega_j/2\pi$ are the frequencies and n_j the quantum numbers. This *harmonic approximation* is the one normally used to specify the lattice vibrations for a particular crystal.

Usually the total internal energy of a crystal is considered to be composed of a lattice energy U (that is, $\Phi_0(x)$) which is not a quadratic function, and a thermal energy, which is the sum of the kinetic and potential energies of the $3n$ oscillators. In this report such theories will be called *harmonic* theories and the term *anharmonic* theory reserved for those cases where the thermal energy has been calculated from an anharmonic potential function. Unfortunately the mathematical difficulties in solving the wave equation for anharmonic potentials ($\Phi_0^{(3)}(u^3)$, etc. not zero) are very great, so that in most cases the anharmonicity has been treated as a small perturbation of the harmonic solution. Recently, however, some progress has been made by Born (1952) and Hooton (1955 a) towards solving the problem of strong anharmonicity in lattice vibrations. Such a theory is essential for lattices in which the thermal energy is large compared with the static potential energy (e.g. solid helium). It approximates to the usual treatment of anharmonicity as a small perturbation when the thermal energy is small.

2.2. *Interatomic Potentials*

The lattice energy U is, in general, a function of the positions of all the nuclei (coordinates x) and hence still a many-body property. It can be reduced to a summation over pairs of lattice points only when the forces between the lattice particles are 'two-body' forces, i.e. independent of the positions of the other lattice particles. When the two-body forces are also 'central' forces, the summation is further simplified and depends only on the magnitude of the vectors between the pairs of particles, i.e. is independent of their relative orientations. The simplification of a central, two-body force, with potential function $\phi(r)$ between lattice particles at distance r apart, is therefore commonly made in calculations of the stability, elasticity and equations of state of crystals. An alternative assumption, that the binding force is non-central but only affects a few of the particles around any lattice point, is sometimes made in calculating the frequencies of vibration and specific heats of metallic crystals. The validity of these assumptions for particular crystals will thus depend on the symmetry and range of the forces between their lattice particles. Although some of these assumptions can be applied to other non-metallic crystals (see Born and Huang 1954, Chapter 1), they are all very nearly true for molecular crystals.

The simplest molecular crystals are the atomic lattices of the inert gas substances, where each atom is bound to its twelve nearest neighbours with the short-range, central, van der Waals force. This has been accurately evaluated by second-order perturbation calculations of the interactions of pairs of atoms. In so far as it is a two-body force in the lattice, it can be summed over all pairs and estimates made of the relative contributions of the dipole–dipole and other van der Waals terms to the cohesive energy.

According to quantum theory the van der Waals potential may be written

$$\phi_{\mathrm{W}}(r) = -\left(\frac{c_1}{r^6}+\frac{c_2}{r^8}+\frac{c_3}{r^{10}}\right) \qquad \ldots\ldots(12)$$

where c_1, c_2, c_3 are constants. Margenau (1939) has reviewed and extended the earlier calculations and quotes the values shown in table 1. Only the first constant c_1 is known accurately, but the ratios c_2/c_1r^2 and c_3/c_1r^4 indicate the relative contributions of the higher terms. These ratios have been evaluated for r equal to r_{m}, the experimental nearest-neighbour distance at 0°K. (An allowance has been made (see § 2.5.4) for the effects of the zero-point energy in each case.)

Table 1. Van der Waals Potentials for Pairs of Inert Gas Atoms

	r_{m} (Å)	$c_1 \times 10^{60}$ (erg cm^6)	$c_2 \times 10^{76}$ (erg cm^8)	$c_3 \times 10^{92}$ (erg cm^{10})	c_2/c_1r^2	c_3/c_1r^4	$c \times 10^{60}$ (erg cm^6)
He	3·0	1·23	1·89	1·65	0·17	0·010	1·46
Ne	3·2	4·67	6·9	5·3	0·14	0·011	5·38
A	3·8	55·4	120	136	0·15	0·012	64·4
Kr	4·0	107	275	370	0·16	0·013	125
Xe	4·4	233	710	1120	0·16	0·013	273

It is apparent from these ratios that the higher terms make up at the most 18% of the van der Waals energy at r_{m}, so that the dipole–dipole interaction is dominant in all the inert gas crystals. Many lattice theories are based on just one attractive term in the potential,

$$\phi_{\mathrm{W}}(r) = c/r^6, \quad \text{where} \quad c = \left(c_1+\frac{c_2}{r_{\mathrm{m}}{}^2}+\frac{c_3}{r_{\mathrm{m}}{}^4}\right) \qquad \ldots\ldots(13)$$

is the ‘ effective ’ van der Waals constant having values shown in the last column of table 1. In some theories two terms $(c_1/r^6+c_2/r^8)$ are used in the attractive potential, but the third term, which never contributes more than 1·3%, has been neglected in all cases.

The repulsive energy has not been evaluated nearly so accurately, although first-order perturbation calculations have been made for He, Ne and A. These show that the repulsion is a two-body force of very short range and that it can be represented by a central interaction potential of the type

$$\phi_r(r) = b\exp(-r/\rho). \qquad \ldots\ldots(14)$$

Theoretical considerations alone therefore lead to a total interatomic potential for the inert gas atoms of the form

$$\phi(r) = \phi_r(r) + \phi_W(r) = b \exp(-r/\rho) - c/r^6 \qquad \ldots\ldots(15)$$

where the constants ρ and c are fairly well established. The exponential form for the repulsive potential makes calculation of the lattice properties rather complicated and, in any case, is perhaps not valid in the region near the minimum of the total potential, which is of course the essential part in considering the properties of lattices. Many authors have therefore used the simple ' model ' potential of Mie (1903)

$$\phi(r) = A/r^n - B/r^m \quad (n > m), \qquad \ldots\ldots(16)$$

which is based on quite general considerations of equilibrium under a short-range repulsion and a longer range attraction. For the inert gases, it is reasonable to put $m = 6$ and Lennard-Jones (1937) showed that the properties of the crystals and gases could be well correlated by putting $n = 12$, that is,

$$\phi(r) = A/r^{12} - B/r^6. \qquad \ldots\ldots(17)$$

This (12, 6) potential has the simplest analytical form and many calculations based on ' model ' potentials make use of it. We shall consider in § 4 how far this and other ' model ' potentials are in agreement with the quantum-theoretical calculations.

The great simplification of using a central force law with a Mie–Lennard-Jones potential is only possible with the inert gas crystals because other molecular crystals (except perhaps CH_4) consist of non-spherical molecules. Even the methane lattice is not close-packed, so that its binding force is not completely of the central type.

A feature of the inert gas crystals is their low cohesive energy E_0, and therefore low melting-points and boiling-points. The zero-point energy ($E_z = 9R\Theta_D/8$, where Θ_D is the Debye temperature) is not negligible for these substances, but, as will be seen from table 2, its magnitude as a fraction of the cohesive energy is relatively small for the heavier elements. Argon is fairly typical of these in having comparatively little zero-point energy.

Table 2. Experimental Cohesive and Zero-point Energies (cal mole^{-1}) for Inert Gas Crystals

	Θ_D	$E_z = 9R\Theta_D/8$	$-E_0$	$E_z/E_0(\%)$
He†	21	47	12	392
Ne	64	140	450	31·1
A	80	180	1850	9·7
Kr	63	140	2590	5·4
Xe	55	120	3830	3·1

† at 25 atmospheres.

We shall use the term *ideal crystal* to refer to the lattices of the four inert gas substances (Ne, A, Kr, Xe) for which—in contrast to helium—the zero-point energy is sufficiently small to justify a weakly anharmonic theory of the vibrational energy.

2.3. *Equilibrium and Stability of Ideal Crystals*

Classically, the lattice formed at 0°K by the condensation of a particular inert gas will depend on its interatomic potential $\phi(r)$ and will be that lattice for which the static potential energy (U_0) is a minimum. Born (1940) has shown that a second condition limits the choice of alternative lattices for atoms with central forces to a few tightly packed ones. This is the condition that a stable lattice must be such that small displacements of the atoms from their lattice sites (for example, by applying a small shearing stress) results in an increase in the lattice energy $U(U > U_0)$. We shall discuss first the Born condition for mechanical stability of a number of lattices under central forces and then consider the question of which lattice has the lowest static potential energy for the ideal crystal.

2.3.1. *Mechanical stability of cubic and hexagonal lattices.*

The simplest crystal lattices are monatomic Bravais lattices, that is, lattices which are built from three basic vectors ($\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$) and which have identical atoms on each lattice point. The lattice points are the corners of the parallelepipeds (unit cells) bounded by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ and are given by

$$\mathbf{r}(l) = l_1\mathbf{a}_1 + l_2\mathbf{a}_2 + l_3\mathbf{a}_3, \qquad \ldots\ldots(18)$$

l_1, l_2, l_3 being integers.

These integers are somewhat limited for the centred lattices of the cubic system, but the addition of a central atom to the simple hexagonal lattice, while not destroying the hexagonal symmetry, does destroy its Bravais nature. The new lattice is not a Bravais lattice, since the basic lattice vectors ($\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$) do not generate all the lattice points, while the position vector does not lead by repetition to another lattice point.

The calculation of the elastic deformation energy $(U - U_0)$ was first made by Born (1940) and Misra (1940) for the cubic Bravais lattices. For a uniform shear of the lattice, the unit cubes are distorted into parallelepipeds, so that the new lattice points are given by $\mathbf{r}(s, l) = \Sigma_\alpha l_\alpha \mathbf{a}_\alpha$, where $\mathbf{a}_\alpha$ are now the vectors forming the (strained) parallelepiped. The displaced lattice points are at distance $r^{s,l}$ from the origin, given by $(r^{s,l})^2 = \Sigma_{\alpha\beta}(\mathbf{a}_\alpha . \mathbf{a}_\beta)\, l_\alpha l_\beta$, where $\alpha, \beta = 1, 2, 3$, represent the rectangular components of the vector. Hence

$$(r^{s,l})^2 = a^2(l_1^2 + l_2^2 + l_3^2) + a^2 \sum_{\alpha\beta} l_\alpha l_\beta e_{\alpha\beta}, \qquad \ldots\ldots(19)$$

where $e_{\alpha\beta} = (\mathbf{a}_\alpha . \mathbf{a}_\beta - a^2\delta_{\alpha\beta})/a^2$ are the strain components of elasticity theory and $\delta_{\alpha\beta}$ is the Krönecker symbol $(= 0, \alpha \neq \beta;\ = 1, \alpha = \beta)$. In (19) the first term is just the square of the distance r^l in the undistorted lattice, and a measure of the distortion is given by

$$\rho^l = \tfrac{1}{2}a^2 \sum_{\alpha\beta} l_\alpha l_\beta e_{\alpha\beta}. \qquad \ldots\ldots(20)$$

In the ideal crystals, the lattice potential U may be obtained by summing over all $N/2$ pairs of lattice points, assuming that each pair interacts with the Mie (n, m) potential of (16), and omitting the origin $(l = 0)$. For the distorted lattice,

$$U = \tfrac{1}{2}N \Sigma\, \phi(r^{s,l}) = \tfrac{1}{2}N \sum_l \phi\{[(r^l)^2 + 2\rho^l]^{1/2}\},$$

may be expanded in terms of the undistorted potentials $\phi^l = \phi(r^l)$, and

$$D = (1/r)(d/dr)$$

to give

$$U = \tfrac{1}{2}N\sum_l \left\{\phi^l + \rho^l D\phi^l + \frac{(\rho^l)^2}{2!}D^2\phi^l + \ldots\right\}. \qquad \ldots\ldots(21)$$

Comparing this with (6) shows that $U_0 = \frac{1}{2}N\Sigma_l \phi^l$ is just the static lattice energy, while $U_1 = \frac{1}{2}N\Sigma_l \rho^l D\phi^l$ will be zero for static equilibrium (cf. equation (9)).

Hence the increase in energy due to the distortion is given, to the first order, by the quadratic term in (21), which (with 20) is

$$U_2 = \tfrac{1}{2}N\sum_l \left\{\frac{a^4}{8}\left(\sum_{\alpha\beta} l_\alpha l_\beta e_{\alpha\beta}\right)^2 D^2\phi^l\right\}.$$

The volume per atom v in each case is ga^3, where $g = 1$ (simple cubic), 4 (body-centred cubic) and 2 (face-centred cubic). The elastic energy density u is therefore

$$u = \frac{U_2}{Nv} = \frac{1}{2g}\sum_l \frac{a}{8}\left(\sum_{\alpha\beta} l_\alpha l_\beta e_{\alpha\beta}\right)^2 D^2\phi^l. \qquad \ldots\ldots(22)$$

In cubic crystals the thirty-six elastic constants (c_{11} to c_{66}), which are given by the generalized Hooke's law, reduce to only three independent ones (c_{11}, c_{12}, c_{44}) and the elastic energy density is given by

$$u = \frac{c_{11}}{8}(e_{11}{}^2 + e_{22}{}^2 + e_{33}{}^2) + \frac{c_{12}}{4}(e_{11}e_{22} + e_{22}e_{33} + e_{33}e_{11}) + \frac{c_{44}}{2}(e_{12}{}^2 + e_{23}{}^2 + e_{31}{}^2). \qquad \ldots\ldots(23)$$

Comparing coefficients between (22) and (23), and allowing for cubic symmetry in the summations,

$$\left.\begin{aligned} c_{11} &= \frac{a}{2g}\sum_l D^2\phi^l l_1{}^4 \\ c_{12} = c_{44} &= \frac{a}{2g}\sum_l D^2\phi^l l_1{}^2 l_2{}^2. \end{aligned}\right\} \qquad \ldots\ldots(24)$$

The equality of c_{12} and c_{44} is the *Cauchy relation* for cubic crystals and is thus satisfied in all cases for central forces in ideal crystals, when we neglect zero-point energy. The condition that the energy density shall be positive reduces (Born 1940) to

$$c_{11} + 2c_{12} > 0, \quad c_{11} - c_{12} > 0, \quad c_{12} = c_{44} > 0. \qquad \ldots\ldots(25)$$

The solution of these conditions, and the evaluation of U_0, requires the calculation of a number of lattice sums for the potential $\phi(r)$ over the undistorted lattice. If we let $r^l/a = L = (l_1{}^2 + l_2{}^2 + l_3{}^2)^{1/2}$, then the sums involved are (Misra 1940)

$$S_p = \sum_l \frac{1}{L^p}, \quad S_p{}^{(1)} = \sum_l \frac{l_1{}^2}{L^p}, \quad S_p{}^{(2)} = \sum_l \frac{l_1{}^4}{L^p}, \quad S_p{}^{(3)} = \sum_l \frac{l_1{}^2 l_2{}^2}{L^p}, \qquad \ldots\ldots(26)$$

each summation being over all neighbours of one lattice point at the origin.

In terms of these lattice sums and the (n, m) Mie potential, the inequalities (25) become

$$n > m, \quad A(n) > A(m), \quad B(n) > B(m),$$

where

$$\left.\begin{aligned} A(n) &= \tfrac{1}{2}(n+2)\left(9\frac{S_{n+4}{}^{(2)}}{S_n} - 1\right) \\ B(n) &= \tfrac{1}{2}(n+2)\left(1 - \frac{3S_{n+4}{}^{(2)}}{S_n}\right). \end{aligned}\right\} \quad \ldots\ldots(27)$$

Thus a particular crystal will be mechanically stable if its lattice structure and interatomic potential lead to functions $A(n)$ and $B(n)$ which are *both* monotonically increasing with n.

The lattice sums S_p were first evaluated by Jones and Ingham (1925) for $4 \leqslant p \leqslant 30$ using 'Epstein's zeta functions'. Misra (1940) used the theta functions for $4 \leqslant p \leqslant 10$, and for $p > 10$ summed partly by an integral, to obtain S_p, $S_p{}^{(1)}$, $S_p{}^{(2)}$, $S_p{}^{(3)}$ for $4 \leqslant p \leqslant 15$. He then plotted $A(p)$ and $B(p)$ for each cubic lattice against p, as shown in figure 1. These curves show that the simple cubic lattice is always unstable, since $A(p)$ is increasing while $B(p)$ is decreasing for all p, and that the face-centred cubic lattice is always stable. The body-centred cubic has $B(p)$ always increasing, but $A(p)$ only increasing for $p < 5$. It is thus stable only when $p < 5$ and, since $n > m$, this means $n < 5$, i.e. the repulsive potential must be even 'softer' than the van der Waals attraction ($m = 6$). Hence only the face-centred, of the three cubic lattices, is mechanically stable for ideal crystals.

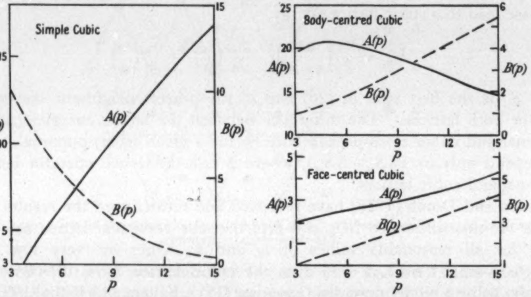


Figure 1. Lattice sums $A(p)$ and $B(p)$ for the simple, body-centred and face-centred cubic lattices. (*After* Misra.)

Born (1942, 1944) has investigated the mechanical stability of the simple and close-packed hexagonal lattices, under central force interactions. He did not completely evaluate the lattice sums, but showed that for a large number of neighbours, the stability conditions were satisfied for the close-packed crystal, but not for the simple hexagonal lattice, which is always unstable. For the ideal crystals, with central force interactions, classical considerations thus show that only the face-centred (close-packed) cubic and close-packed hexagonal lattices are mechanically stable structures.

2.3.2. *Absolute stability of close-packed lattices.*

The actual structure of a crystal at any temperature and pressure will be that structure for which the Gibbs free energy

$$G = E + pV - TS \qquad \ldots\ldots(28)$$

is a minimum. For ideal crystals at zero pressure and absolute zero of temperature, the only term in the Gibbs free energy is the static lattice energy U_0, and so the most stable structure is the one for which

$$U_0 = \tfrac{1}{2}N \sum_l \phi^l \qquad \ldots\ldots(29)$$

is a minimum. The Mie (n, m) potential can conveniently be written

$$\phi^l = u\left(\frac{nm}{n-m}\right)\left\{-\frac{1}{m}\left(\frac{r_0}{r^l}\right)^m + \frac{1}{n}\left(\frac{r_0}{r^l}\right)^n\right\}, \quad n > m, \qquad \ldots\ldots(30)$$

where u is the minimum potential energy for a pair of atoms at the equilibrium separation r_0. The summations over the allowed values of l will be characteristic of the particular close-packed lattice and of the particular potential (n, m), and in each case lead to a static lattice energy,

$$U_0 = \frac{Nu}{2}\left(\frac{nm}{n-m}\right)\left\{-\frac{S_m}{m}\left(\frac{r_0}{a}\right)^m + \frac{S_n}{n}\left(\frac{r_0}{a}\right)^n\right\} \qquad \ldots\ldots(31)$$

where S_p is the first sum in (26) and a, the nearest neighbour distance, is the same for both lattices. The difference between the lattice energies (ΔU_0) of the hexagonal and cubic close-packed lattices, for a given (n, m) potential, will therefore depend only on $(\Delta S_n - \Delta S_m)$, where Δ is a difference operator between the hexagonal and cubic lattices.

Barron and Domb (1955) have reviewed and recalculated the results of several authors for this difference ΔU_0 and find that the *hexagonal* lattice has the lower energy for all reasonable values of m and n. Only for very low values of n ($n = 6, m = 5; 5 \leqslant n \leqslant 8, m = 4$) does the cubic lattice have the lower energy. Similarly, using a model potential (equation (15)), Kihara and Koba (1952) showed that the hexagonal lattice was always more stable, except for very soft repulsive constants. But such wide potential wells are not reasonable approximations for van der Waals crystals ($n \gg m = 6$), nor do they lead to equations of state that fit the experimental data for the inert gases or their crystals. Barron and Domb (1955) show further, for a Lennard-Jones (12, 6) potential, that the lattice energy difference ΔU_0 is only 0·01% of U_0 at zero pressure and vanishes at a high pressure (10^5 atm for A). Above this transition pressure, the cubic lattice is the more stable one. As is well known, however, all the inert gas crystals (except helium, which is non-ideal in any case) crystallize in the *cubic* close-packed lattice at low pressure.

A possible explanation for this discrepancy is that the direct summation of a central, two-body potential of the Mie–Lennard-Jones type, while a very good approximation to the van der Waals interactions for a pair of inert gas atoms, is not an accurate representation of the many-body problem of the ideal crystal lattice. There may, in fact, be small amounts of energy due to non-central, many-body interactions, which could change U_0 sufficiently to make the cubic lattice the stable one at zero pressure and temperature. Axilrod and Teller (1943) first discussed interactions of the dipole–dipole type between three atoms, and showed that their contribution to the interatomic potential $\phi_A(r)$ depended on the shape of the triangle formed by the three atoms. In general, this dipole–dipole–dipole (or triple–dipole) interaction leads to a potential (cf. equation (12))

$$\phi_A(r) = c_4/r^9, \qquad \ldots\ldots(32)$$

where $\phi_A(r)/\phi_W(r)$ is about $+1\%$ for a collinear array of three inert gas atoms and about -3% for an equilateral triangle.

In a later paper, Axilrod (1951) summed this triple–dipole interaction over a cylinder of the lattice, which included 36 atoms in each of 9 close-packed layers. The layers were arranged according to each type of close-packed lattice, that is, ABCABC (figure 2) for the cubic, and ABABAB for the hexagonal close-packed.

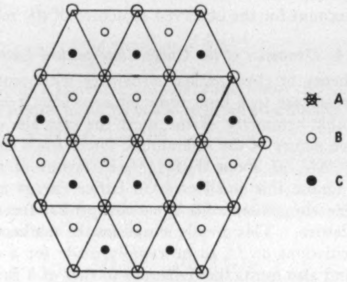


Figure 2. Atomic layers in the cubic close-packed (ABCABC) and hexagonal close-packed (ABABAB) lattices. (*After* Axilrod.)

Altogether 50 000 terms were summed for each case and the lattice energies U_A due to $\phi_A(r)$ calculated in each case. They were both found to be *repulsive* energies, that for the hexagonal lattice being larger by 0·06% of U_A than that for the cubic lattice. However, U_A is only a small contribution to the total lattice energy U_0, and was shown by Axilrod to be about 2% of U_0 for neon, rising to 9% for xenon. Thus the difference ΔU_A in the most favourable case (xenon) is only $-0{\cdot}005\%$ of U_0 and so is not quite sufficient in itself to counteract the $+0{\cdot}01\%$ difference due to the ordinary central-force interactions.

Similar calculations have not yet been made for the non-additive contributions to the repulsive energy, which is usually approximated by a central force with a two-body potential given by $\phi_r(r)$ of equation (14). But the repulsive interaction between a pair of nearest neighbours must be slightly influenced by the presence of the other nearest neighbours (as has been emphasized by Axilrod 1951 and by Prins *et al.* 1952), so that the different spatial arrangements of even the 12 *nearest* neighbours in the hexagonal and cubic lattices would give a repulsive energy difference between them. Barron and Domb (1955) point out that the triple–dipole difference (ΔU_A) becomes steadily less important for the lighter atoms, so that if the repulsive energy difference was similar in its behaviour, the two non-additive contributions might account not only for the cubic symmetry of ideal crystals, but also for the observed hexagonal symmetry of the lightest inert gas crystal, solid helium.

On the other hand, the very large zero-point energy in solid helium might itself be sufficiently different in the two lattices to account for the observed structure. Barron and Domb (1955) have shown that the small amount of zero-point energy present in ideal crystals is not sufficiently different for the two close-packed lattices to account for their observed structure. The difference ($\Delta E_z/U_0$) varies from $-0{\cdot}003\%$ for neon to $-0{\cdot}0003\%$ for xenon and favours the cubic lattice. Jansen and Dawson (1955) have also calculated the effect of zero-point energy on the lattice stability of ideal crystals. They found the change in the lattice constant at equilibrium due to the additional energy and hence found the internal energy difference ($\Delta E = \Delta U_0 + \Delta E_z$) for the two close-packed lattices. They agree with Barron and Domb that only non-additive contributions to the lattice energy can account for the observed structure of the inert gas crystals.

2.4. *Dynamics of the Cubic Close-packed Lattice*

Although the theory of close-packed structures with central forces has been worked out with great care to a high order of accuracy, it has failed to account for the observed cubic symmetry of the inert gas crystals. The assumption is that the non-additive nature of the interatomic forces leads to a difference in the static lattice energy ΔU_0 of about $0{\cdot}01\%$ U_0 in favour of the cubic structure. However, we shall ignore this small error in lattice energy and assume that the latter is given by the summation of a two-body central force over the observed cubic close-packed lattice. This greatly simplifies the mathematical development, since it enables calculations on U_0 to be rapidly made for a selected interatomic potential function and also limits the dynamics to that of a Bravais lattice.

The N lattice particles of a Bravais lattice in equilibrium are at positions given by $\mathbf{r}(l)$ with components x_α^l. When slightly displaced from equilibrium the components of the displacement vector $\mathbf{u}(l)$ are u_α^l for the particle l and $u_\beta^{l'}$ for its neighbour l', where $l(l_1 l_2 l_3)$ and $l'(l_1' l_2' l_3')$ enumerate the lattice sites in the unstrained crystal. Under the assumption of additive, central forces the potential energy between the particles l, l' will depend only on their distance apart, $r^{l-l'}$, and so will be denoted by $\phi^{l-l'}$. In the harmonic approximation (§ 2.1) the total lattice potential U for the dynamic lattice is given by a Taylor expansion in the particle displacements up to the quadratic term (cf. equation (21))

$$U = U_0 + U_1 + U_2. \qquad \ldots\ldots(33)$$

Here $U_0 = \frac{1}{2} N \Sigma_{ll'} \phi^{l-l'}$ is the static lattice potential,

$$U_1 = \sum_{ll'} \sum_{\alpha} \left(\frac{\partial U}{\partial u_\alpha^l} \right)_0 u_\alpha^{l-l'}$$

is the first-order term and

$$U_2 = \tfrac{1}{2} \sum_{ll'} \sum_{\alpha\beta} \left(\frac{\partial^2 U}{\partial u_\alpha^l \partial u_\beta^{l'}} \right)_0 u_\alpha^{l-l'} u_\beta^{l'-l'}$$

is the second-order term in the total potential, each derivative referring to the equilibrium lattice.

In the dynamic lattice, the particles that have been displaced by arbitrary amounts $\mathbf{u}(l)$ are acted on by quasi-elastic restoring forces (Born and von Kármán 1913) according to the Newtonian equation

$$m\ddot{u}_\alpha^{\,l} = -\partial U/\partial u_\alpha^{\,l}, \qquad \ldots\ldots(34)$$

where the derivative is now evaluated for the nuclei in their displaced positions. Using the Taylor expansion for U (equation (33)),

$$m\ddot{u}_\alpha^{\,l} = -\left(\frac{\partial U}{\partial u_\alpha^{\,l}}\right)_0 - \sum_{l'\beta}\left(\frac{\partial^2 U}{\partial u_\alpha^{\,l}\,\partial u_\beta^{\,l'}}\right)_0 u_\beta^{\,l-l'}. \qquad \ldots\ldots(35)$$

Here the term $(\partial U/\partial u_\alpha^{\,l})_0$ represents the α-component of the force on the particles in the equilibrium configuration and so must vanish (cf. equation (9)). The force constants in the second sum (following Born) may be conveniently written $U_{\alpha\beta}^{\,l-l'}$, so that (35) becomes

$$m\ddot{u}_\alpha^{\,l} = -\sum_{l'\beta} U_{\alpha\beta}^{\,l-l'}\,u_\beta^{\,l-l'}. \qquad \ldots\ldots(36)$$

This equation represents a system of simultaneous, linear differential equations, and can be written more simply in terms of the reduced coordinates, $D_{\alpha\beta}^{\,l-l'} = m^{-1}U_{\alpha\beta}^{\,l-l'}$ and $w_\alpha^{\,l} = m^{1/2}u_\alpha^{\,l}$, $w_\beta^{\,l-l'} = m^{1/2}u_\beta^{\,l-l'}$, giving

$$\ddot{w}_\alpha^{\,l} = -\sum_{l'\beta} D_{\alpha\beta}^{\,l-l'}\,w_\beta^{\,l-l'}. \qquad \ldots\ldots(37)$$

In (37) the mass is omitted explicitly and the term $D_{\alpha\beta}^{\,l-l'}$ is the *dynamical matrix* of the lattice, whose *secular determinant*

$$|D_{\alpha\beta}^{\,l-l'} - \omega^2\delta_{\alpha\beta}| = 0 \qquad \ldots\ldots(38)$$

is a cubic equation for the squares of the angular frequencies ω in the harmonic vibrations $w_\alpha^{\,l} = W_\alpha^{\,l}\exp(-i\omega t)$, where $W_\alpha^{\,l}$ are constants.

The dynamical matrix can also be transformed into the reciprocal lattice formed by the reciprocal basic vectors

$$\mathbf{b}_1 = \frac{\mathbf{a}_2\times\mathbf{a}_3}{v}, \quad \mathbf{b}_2 = \frac{\mathbf{a}_3\times\mathbf{a}_1}{v}, \quad \mathbf{b}_3 = \frac{\mathbf{a}_1\times\mathbf{a}_2}{v}, \qquad \ldots\ldots(39)$$

where v is the volume of the cell defined by the basic lattice vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. A reciprocal lattice vector is $\mathbf{q}(h) = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$, where $h(h_1, h_2, h_3)$ are integers, and its product with the general lattice vector $\mathbf{r}(l)$ is also an integer

$$\mathbf{r}(l)\,.\,\mathbf{q}(h) = h_1 l_1 + h_2 l_2 + h_3 l_3 = (hl). \qquad \ldots\ldots(40)$$

Hence the expression $\exp\{2\pi i\mathbf{q}\,.\,\mathbf{r}(l)\}$, where $\mathbf{q}$ is now *any* point in reciprocal space is periodic in the reciprocal lattice. An alternative set of solutions of (37) is therefore given by

$$w_\alpha^{\,l} = V_\alpha\exp\{2\pi i\mathbf{q}\,.\,\mathbf{r}(l)\}\exp(-i\omega t), \qquad \ldots\ldots(41)$$

where V_α are constants. For each point $\mathbf{q}$, the frequencies are determined by

$$|C_{\alpha\beta}(\mathbf{q}) - \omega^2\delta_{\alpha\beta}| = 0, \qquad \ldots\ldots(42)$$

where $C_{\alpha\beta}(\mathbf{q}) = \Sigma_l D_{\alpha\beta}{}^l \exp\{-2\pi i\mathbf{q}.\mathbf{r}(l)\}$ is the transform of $D_{\alpha\beta}{}^l$. These solutions are plane waves of wave number equal to $|\mathbf{q}|$ and wave normal parallel to $\mathbf{q}$.

The *normal modes* of vibration of the lattice are linear combinations of solutions (41) of the same frequency that satisfy the conditions at the surface of the crystal. For macroscopic specimens, the distribution of normal modes over frequency is independent of the shape of the specimen or of particular boundary conditions imposed (Ledermann 1944, Peierls 1954). The simplest method is therefore to use the *periodic boundary condition* (Born 1923, p. 557), in which the specimen is a large cube of volume V such that corresponding points on opposite faces move in exactly the same way. Then the normal modes are those solutions (41) which have wave numbers given by $\mathbf{q} = V^{-1/3}(n_1, n_2, n_3)$, where n_1, n_2, n_3 are integers. In reciprocal space, these are a set of points of uniform density V and, since $\mathbf{q}(h).\mathbf{r}(l)$ is an integer, there is no loss of generality if we consider these points for just one cell of the reciprocal lattice, volume $= |\mathbf{b}_1.\mathbf{b}_2 \times \mathbf{b}_3| = 1/v$. Hence the total number of points is $V/v = N$, where N is the number of lattice cells in the crystal volume V. In a Bravais lattice, there are three independent modes for each wave vector $\mathbf{q}$ and so altogether there are $3N$ vibrational modes of frequency $\omega_j/2\pi$ ($j = 1, 2, 3$), given by (42). Thus equations (41) and (42) provide a complete solution of the lattice vibrations and their frequency distribution.

For the cubic close-packed lattice, the reciprocal lattice is a body-centred cubic. The symmetry relations for any cubic lattice further limit the number of points in the reciprocal cell for which solutions of the secular determinant (42) need be evaluated to 1/48th of the total cell. In practice, the simplifying assumption that only the nearest and next nearest neighbours in the lattice interact with one another is normally made in calculating a particular frequency spectrum. This simplifies the secular determinant (42) of $C_{\alpha\beta}(\mathbf{q})$, which becomes

$$\begin{vmatrix} A(xyz)-m\omega^2 & B(xy) & B(xz) \\ B(yx) & A(yzx)-m\omega^2 & B(yz) \\ B(zx) & B(zy) & A(zxy)-m\omega^2 \end{vmatrix} = 0, \qquad \ldots\ldots(43)$$

where the diagonal terms A and non-diagonal terms B can be further simplified for each particular lattice. If $\boldsymbol{\alpha}$[†] is the restoring force constant between adjacent atoms when the motion is in the line connecting them and $\boldsymbol{\gamma}$ that when the motion is normal to the same line, then the determinant elements for the cubic close-packed lattice are (Leighton 1948):

$$A(xyz) = 2\boldsymbol{\alpha}(2-\cos x\cos y - \cos x\cos z) + 4\boldsymbol{\gamma}\sin^2 x,$$

$$B(xy) \;\; = 2\boldsymbol{\alpha}\sin x\sin y.$$

The full frequency spectrum of this lattice has been found by Leighton by modelling the constant frequency surfaces of the secular determinant in plaster of Paris and measuring the volume enclosed between successive surfaces for 1/48th of the reciprocal cell. The spectrum was evaluated for $\boldsymbol{\gamma}/\boldsymbol{\alpha} = 0$ and $0{\cdot}1$ and is shown in figure 3(*a*), where $\lambda^2 = m\omega^2/2\boldsymbol{\alpha}$ and I, II, III are the three acoustic modes whose sum is the distribution $G(\lambda)$. This total spectrum is compared with that of the Debye $[G(\nu) \propto \nu^2, \nu < \nu_D]$ and Einstein (ν_E) models in figure 3(*b*). It is obvious that the Debye model gives the better approximation to the true frequency spectrum, but that it is not exact, as had been shown for other cubic lattices by Blackman (1941). Following Blackman it has become customary to compare the frequency spectra of lattices with that of Debye by calculating the equivalent Debye temperature (Θ_D) as a function of temperature for each lattice. Such 'Debye plots' for Leighton's spectra are shown in § 3.3.1, where the parameter

---

[†] We use $\boldsymbol{\alpha}, \boldsymbol{\gamma}$ to distinguish the force constants from the expansivity α and Grüneisen's constant γ.

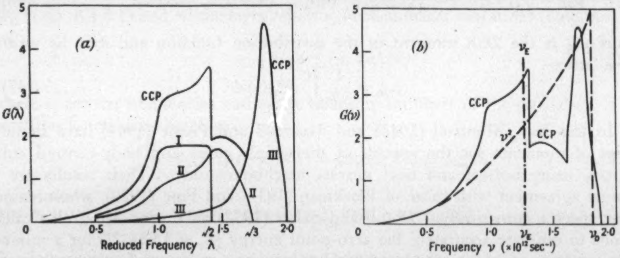


Figure 3. Frequency spectra for the cubic close-packed lattice $(\gamma/\alpha = 0)$. (*a*) The three separate acoustic modes I, II, III and their sum (CCP); (*b*) the sum (CCP) compared with the equivalent Debye (ν^2, ν_D) and Einstein (ν_E) models. (*After* Leighton.)

$$\beta = \frac{h}{\pi k}\sqrt{\frac{\alpha}{2m}}. \qquad \ldots\ldots(44)$$

They show that Θ_D falls at low temperatures, but for $T > \Theta_D/4$, say, it becomes constant up to high temperatures. This low temperature variation of Θ_D corresponds to the rise of $G(\nu)$ (figure 3(*b*)) above the Debye spectrum at low frequencies.

A different method of evaluating the frequency spectrum was used by Montroll (1942), who expressed the frequency distribution as a sum of the even moments, which he found by an analytical method due to Thirring (1913) without solving the secular determinant for each frequency. If $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N$ are the N values of $\mathbf{q}$ in the reciprocal cell giving the $3N$ frequencies $\nu_j(\mathbf{q}_l)$ $(j = 1, 2, 3;\ l = 1, 2, \ldots, N)$, then the squares of their angular frequencies $\omega_j{}^2(\mathbf{q}_l)$ are the characteristic roots of the $N \times N$ diagonal matrix C, where

$$C = \begin{bmatrix} C(\mathbf{q}_1) & 0 & \ldots & 0 \\ 0 & C(\mathbf{q}_2) & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \ldots & C(\mathbf{q}_N) \end{bmatrix} \qquad \ldots\ldots(45)$$

and where each diagonal element $C(\mathbf{q}_l)$ is a 3×3 matrix $C_{\alpha\beta}(\mathbf{q}_l)$. But the sum of the kth powers of all the characteristic roots of any matrix C is given by the trace (or diagonal sum) of the matrix C^k. Hence the average value of the kth power of the characteristic roots $\nu_j^2(\mathbf{q}_l)$ of $C/4\pi^2$ is μ_{2k} where

$$\mu_0 = \frac{1}{3N}\mathrm{Tr}[C^0] = 1,$$

$$\mu_{2k} = \frac{1}{3N}\mathrm{Tr}[C^k] = \frac{1}{3N}\sum_j \sum_l [\nu_j^2(\mathbf{q}_l)]^k, \quad (k = 1, 2, \ldots, \text{etc.}). \qquad \ldots\ldots(46)$$

Thus μ_{2k} is the $2k$th moment of the distribution function and may be written for a lattice

$$\mu_{2k} = \frac{1}{3N}\int_0^\infty \nu^{2k} g(\nu)\, d\nu. \qquad \ldots\ldots(47)$$

In this way, Montroll (1943) and Montroll and Peasle (1944) have found a series of moments for the spectra of the simple cubic and body-centred cubic lattices, using nearest and next nearest neighbours only. Their results are in general agreement with those of Blackman (1937) and Fine (1939), who obtained these spectra numerically. Domb and Salter (1952) have used Montroll's calculations to estimate accurately the zero-point energy ($E_z = 3Nh\mu_1/2$) for a number of cubic lattices. In particular they found μ_1 from a series of values of μ_{2k} for the cubic close-packed lattice with nearest neighbour interactions only. Two independent methods of calculation gave $E_z/Nh\nu_{\mathrm{m}} = 1\cdot0225$ and $1\cdot0220$, where ν_{m} is the true maximum frequency of the lattice (given by $\lambda = 2$ in figure 3(a)) and is equal to $1/\pi(2\alpha/m)^{1/2}$. These values of E_z may be compared with the Debye model value,

$$E_z = \frac{9N}{\nu_{\mathrm{D}}^3}\int_0^{\nu_{\mathrm{D}}} \tfrac{1}{2}h\nu\,.\,\nu^2\, d\nu = \tfrac{9}{8}Nh\nu_{\mathrm{D}} = 1\cdot125Nh\nu_{\mathrm{D}}, \qquad \ldots\ldots(48)$$

which is about 10% larger if $\nu_{\mathrm{D}} = \nu_{\mathrm{m}}$. In terms of the Debye temperature ($\Theta_{\mathrm{D}} = h\nu_{\mathrm{D}}/k$), E_z(Debye) $= \tfrac{9}{8}R\Theta_{\mathrm{D}}$. Domb and Salter show that this expression is only accurate if Θ_∞, the limiting value of Θ as $T \to \infty$, is used for Θ_{D}, giving $E_z/Nh\nu_{\mathrm{m}} = 1\cdot0270$, in error by only ½% from the lattice value. This has the additional advantage that Θ_∞ can be found directly from the second moment of the lattice spectrum, i.e.

$$\Theta_\infty = \frac{h}{k}\sqrt{\frac{5\mu_2}{3}} = \frac{0\cdot9129}{\pi}\frac{h}{k}\sqrt{\frac{2\alpha}{m}}, \qquad \ldots\ldots(49)$$

where the numerical factor applies to all three cubic lattices. A refined calculation of Θ_∞ allowing for the effect of all neighbours with a (12, 6) potential showed that the simple formula (49) was accurate to 3% for the close-packed lattice. For this lattice the zero-point energy

$$E_z = \tfrac{9}{8}R\Theta_\infty \qquad \ldots\ldots(50)$$

can therefore be calculated quite accurately.

2.5. *Thermodynamics of Ideal Crystals*

2.5.1. *General equation of state.*

The macroscopic (thermodynamic) properties of crystals (expansivity, specific heat, compressibility, etc.) are thermal averages of the microscopic (dynamic) motions of their lattice particles. The statistical mechanics of such motions shows that the (Helmholtz) free energy F of the system is related to the eigenvalues of the motions through the partition function Z by

$$F = -kT \ln Z. \qquad (51)$$

The total eigenvalue ϵ_t for the ideal crystals is the sum of the eigenvalue U_0 of the total static energy when the n nuclei are fixed in the equilibrium lattice ($\Phi_0^{(0)}(x_0)$ in § 2.1) and all the eigenvalues ϵ_j of the nuclear vibrations (equation 11), i.e.

$$\epsilon_t = U_0 + \sum_{j=1}^{3n} (n_j + \tfrac{1}{2})\, h\nu_j \qquad \ldots\ldots(52)$$

where ν_j are the frequencies and n_j the quantum numbers of the oscillators. The partition function is therefore

$$Z = \sum_{n_j} \exp(-\epsilon_t/kT) = \exp(-U_0/kT) \sum_j Z_j, \qquad \ldots\ldots(53)$$

where
$$Z_j = \exp(-\tfrac{1}{2}h\nu_j/kT) \sum_{n_j=0}^{\infty} \exp(-n_j h\nu_j/kT).$$

Summing over all quantum numbers and expressing the equation in terms of

$$\beta_j = \frac{h\nu_j}{kT}, \quad Z_j = \frac{\exp(-\tfrac{1}{2}\beta_j)}{1-\exp(-\beta_j)}. \qquad \ldots\ldots(54)$$

Therefore, substituting (54) in (53), and using (51), the free energy is

$$F = U_0 + kT \sum_j \tfrac{1}{2}\beta_j + kT \sum_j \ln[1-\exp(-\beta_j)]. \qquad \ldots\ldots(55)$$

Here, following § 2.1, the lattice energy U_0 is a function only of the interatomic distances, or of the volume, in the adiabatic approximation, and the frequencies ν_j are similarly only functions of the volume when found (as in § 2.4) from the harmonic approximation. From the free energy F the internal energy E and the entropy S are given by the thermodynamic relations

$$E = F + TS = F - T \left.\frac{\partial F}{\partial T}\right|_V.$$

Hence
$$E = U_0 + E_z + E_T = U_0 + E_{\mathrm{vib}}, \qquad \ldots\ldots(56)$$

where
$$E_z = \tfrac{1}{2}\sum_j h\nu_j, \quad E_T = kT \sum_j \frac{\beta_j}{\exp(\beta_j)-1}. \qquad \ldots\ldots(57)$$

The specific heat, $C_v = (\partial E/\partial T)_V = (\partial E_T/\partial T)_V$, is seen to depend solely on the frequency spectrum of the lattice (in the harmonic approximation) and so provides a direct way of studying this spectrum experimentally at those low temperatures where only some of the j states are excited. The other thermodynamic properties (including C_p when this is not equal to C_v within experimental error) are derivatives of the free energy F and so depend (56) on both the lattice energy U_0 and the total vibrational energy E_{vib}.

They can therefore only be obtained from the theory under the assumptions of a particular interatomic potential ϕ (§ 2.2) or an arbitrary lattice potential U_0

(e.g. $U_0 = a+b(V-V_0)+c(V-V_0)^2+\dots$ (Slater 1939)). In particular, the equation of state gives the pressure ($p = (-\partial F/\partial V)_T$) as a function of U_0 and E_{vib}, since from (55),

$$p = -\frac{dU_0}{dV} - \sum_j \left(\frac{1}{2}+\frac{1}{\exp(\beta_j)-1}\right)\frac{dh\nu_j}{dV}. \qquad \text{......(58)}$$

The variation of the frequency spectrum with volume is conveniently expressed in terms of the dimensionless parameters γ_j defined by

$$\gamma_j = -\frac{d\ln\nu_j}{d\ln V} = -\frac{V}{\nu_j}\frac{d\nu_j}{dV}. \qquad \text{......(59)}$$

Using these parameters, the general equation of state becomes

$$p+\frac{dU_0}{dV} = \frac{1}{V}\sum_j \gamma_j\left(\frac{1}{2}h\nu_j+\frac{h\nu_j}{\exp(\beta_j)-1}\right) \qquad \text{......(60)}$$

or

$$pV+VU_0'(V) = \sum_j \gamma_j E_{\text{vib}}. \qquad \text{......(61)}$$

The Einstein (1907) and Debye (1912) models of the crystal lattice (§ 2) both provide simple expressions for $\gamma_j E_{\text{vib}}$ (61) in terms of a characteristic temperature (Θ_{E} or Θ_{D}) and a parameter ($\gamma_{\text{E}}, \gamma_{\text{D}}$). In the Einstein theory,

$$\frac{h\nu_j}{k} = \frac{h\nu_{\text{E}}}{k} = \Theta_{\text{E}};\ -\frac{d\ln\nu_j}{d\ln V} = -\frac{d\ln\nu_{\text{E}}}{d\ln V} = \gamma_{\text{E}} \qquad \text{......(62)}$$

and so, from (61),

$$pV+VU_0'(V) = \gamma_{\text{E}} f_{\text{E}}(T/\Theta_{\text{E}}). \qquad \text{......(63)}$$

Similarly, in Debye's theory,

$$\left.\begin{aligned} h\nu_{\text{D}} = k\Theta_{\text{D}};\ -\frac{d\ln\nu_{\text{D}}}{d\ln V} = \gamma_{\text{D}} \\ \text{and so} \quad pV+VU_0'(V) = \gamma_{\text{D}} f_{\text{D}}(T/\Theta_{\text{D}}). \end{aligned}\right\} \qquad \text{......(64)}$$

Here the particular functions f_{E} and f_{D} are characteristic of the single frequency and ' limited continuum ' spectra of the models. On the other hand, Grüneisen (1926) simplified the equation of state (61) to

$$pV+VU_0'(V) = \gamma E_{\text{vib}} \qquad \text{......(65)}$$

which can be obtained from (61) by supposing that all the γ_j are equal, independent of the frequency spectrum. In this case, from (65),

$$V\left.\frac{\partial p}{\partial T}\right|_V = \gamma\left.\frac{\partial E_{\text{vib}}}{\partial T}\right|_V = \gamma C_{\text{v}}$$

and

$$\left.\frac{\partial p}{\partial T}\right|_V = -\left.\frac{\partial p}{\partial V}\right|_T \left.\frac{\partial V}{\partial T}\right|_p = \frac{\alpha}{\kappa_T},$$

where α is the expansivity $[(1/V)(\partial V/\partial T)_p]$ and κ_T the isothermal compressibility $[-(1/V)(\partial V/\partial p)_T]$ of the crystal. Hence the Grüneisen parameter

$$\gamma = \alpha V/\kappa_T C_{\text{v}} = \alpha V/\kappa_S C_{\text{p}}, \qquad \text{......(66)}$$

may be obtained experimentally from these quantities. ' Grüneisen's law ' states that γ is a constant, independent of the volume, so that α is proportional to C_{p} at temperatures low enough for the (smaller) variations in V/κ_S to be neglected.

More generally, γ can be considered as a slowly varying function of V, T characteristic of a particular crystal (Barron 1955). In this case, the values of γ obtained from (65) and (66) are not necessarily identical.

2.5.2. *Law of corresponding states.*

The general equation of state (61) will lead to a law of corresponding states, and hence a reduced equation, for a class of crystals that fulfil certain criteria. Originally, van der Waals on classical grounds supposed that the equation of state for fluids could be written $p/p_c = f(V/V_c, T/T_c)$, where the values of the pressure (p_c), molar volume (V_c) and temperature (T_c) at the critical point were used to form the reduced equation. A number of authors (e.g. de Boer and Michels 1938, Born 1939, Pitzer 1939, Guggenheim 1945) have extended van der Waals' idea to molecular fluids and crystals by expressing the thermodynamic variables (p, V, T) in terms of molecular units. We shall use the notation of de Boer and Michels.

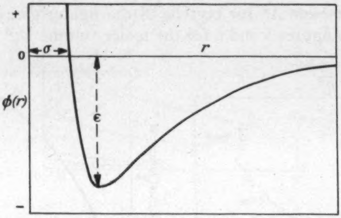


Figure 4. Lennard-Jones (12, 6) interatomic potential function.

In classical statistics a law of corresponding states can be derived, provided that (*a*) the lattice potential U_0 is a sum over central force, two-body interactions $[U = \Sigma\phi(r)]$ and (*b*) the intermolecular potential $\phi(r)$ is of a similar form for all substances. The second condition can be written

$$\phi(r) = \epsilon f(r/\sigma), \qquad \ldots\ldots(67)$$

where ϵ, σ are a characteristic energy and length for a particular substance and the function f is the same for all substances obeying the law. A mathematically simple form of f is the Lennard-Jones (12, 6) potential (figure 4), which as expressed by de Boer, is

$$\phi(r) = 4\epsilon\{(\sigma/r)^{12} - (\sigma/r)^{6}\}. \qquad \ldots\ldots(68)$$

The molecular units for p, V, T are then $\epsilon/\sigma^3, N\sigma^3$ and ϵ/k respectively, and the reduced quantities (p^*, V^*, T^*) are therefore defined as

$$p^* = p\sigma^3/\epsilon,\ V^* = V/N\sigma^3,\ T^* = kT/\epsilon. \qquad \ldots\ldots(69)$$

In terms of these quantities, the classical equation of state may be written in the reduced form

$$p^* = f(V^*, T^*) \qquad \ldots\ldots(70)$$

which will apply therefore to the ideal crystals.

The deviations from this law that are directly attributable to quantum effects have been studied by de Boer (1948) in terms of the de Broglie wavelength, $\Lambda = h/(m\epsilon)^{1/2}$, for the relative motion of two molecules with energy ϵ. In reduced form this wavelength is $\Lambda^* = \Lambda/\sigma$ and appears as a parameter in the reduced wave equation for the nuclear motions. Consequently, in quantum theory, the reduced partition function is $Z^*(V^*, T^*, \Lambda^*)$ and the equation of state becomes

$$p^* = \mathscr{F}(V^*, T^*, \Lambda^*). \qquad \ldots\ldots(71)$$

Thus the law of corresponding states (70) only holds under the conditions for which classical statistics are valid, namely, $\Lambda^* \to 0$.

2.5.3. *Zero-point effects.*

The influence of the zero-point energy E_z (57) on the anharmonicity of the nuclear vibrations in solid helium has already been mentioned (§ 2.2). The general effect of quantum deviations from classical theory can now be considered in terms of de Boer's molecular units. In particular, the striking effect of the large quantum parameters Λ^* for crystals of the lighter gases (H_2, He and their isotopes) is shown in figures 5 and 6 for the molar volume V_0^* and internal energy E_0^* at 0°K. (The experimental values used are listed in the Appendix.) De Boer and Lunbeck (1948) first obtained the Λ^* value for ^{3}He (3·05) from that of ^{4}He by multiplying by $(4/3)^{1/2}$, since the values of ϵ, σ are the same for isotopes. They were then able to extrapolate the graphs of V_0^* and E_0^* (as shown) to find the molar volume ($V_0 = 27$ cm^3) and the internal energy ($E_0 = 5{\cdot}1$ cal) for ^{3}He before it was measured. Recent experiments (Abraham *et al.* 1950, 1952, 1953, Kerr 1954) have confirmed these predictions and give $V_0 = 32$ cm^3, $E_0 = 4{\cdot}5$ cal (under 29 atm pressure). The general effect of the zero-point vibrations is thus an expansion of the lattice and a reduction of the cohesive energy, the results being most marked in the lightest elements, hydrogen and helium.

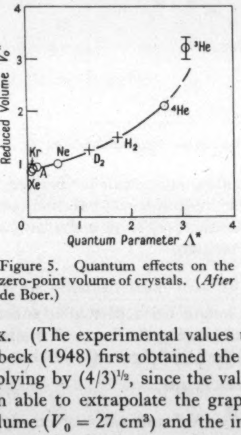


Figure 5. Quantum effects on the zero-point volume of crystals. (*After* de Boer.)

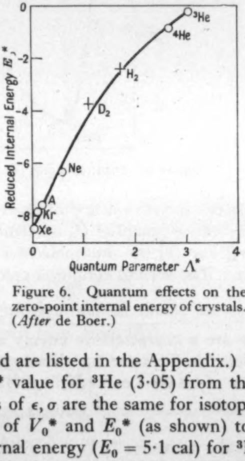


Figure 6. Quantum effects on the zero-point internal energy of crystals. (*After* de Boer.)

Even in the heavier elements like argon and krypton, the zero-point energy has a measurable effect on the bulk properties at 0°K. This was brought out clearly by Salter (1954 a) who used the Domb and Salter expression for the zero-point energy (50) to calculate the elastic constants c_{ij} of the ideal crystals. This calculation showed that, even on the basis of central forces in a harmonic lattice, the zero-point energy destroyed the Cauchy relation ($c_{12} = c_{44}$) for these solids. The equations (50) and (49) give $E_z = 9Nk\Theta_\infty/8 = 9Nh(5\mu_2/3)^{1/2}/8$, where from (46), $\mu_2 = \mathrm{Tr}[C]/3N$. Remembering that each matrix $C_{\alpha\beta}$ is a transform of $D_{\alpha\beta}{}^{l-l'} = m^{-1}U_{\alpha\beta}{}^{l-l'}$, where $U_{\alpha\beta}{}^{l-l'} = (\partial^2 U/\partial u_\alpha{}^l \partial u_\beta{}^{l'})_0$, we obtain

$$\mathrm{Tr}[C] = \frac{N}{4\pi^2 m}\sum_\alpha \sum_l U_{\alpha\alpha}{}^l \quad \text{and} \quad \mu_2 = \frac{1}{12\pi^2 m}\sum_l \nabla^2\phi^l,$$

so that

$$\frac{E_z}{N} = \frac{3h}{16\pi}\left(\frac{5}{m}\sum_l \nabla^2\phi^l\right)^{1/2} = \frac{3\hbar}{8}\left(\frac{5S}{m}\right)^{1/2}, \text{ say.} \qquad \ldots\ldots(72)$$

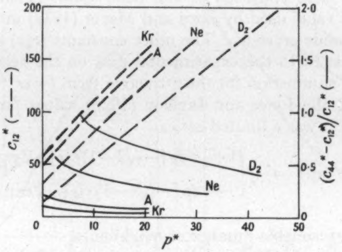


Figure 7. Quantum effects on the zero-point elastic constants of crystals as a function of pressure (in reduced units). (*After* Salter.)

The total internal energy of the lattice at 0°K ($E_0 = U_0 + E_z$) can then be expanded in terms of the strain components, as was done by Born and Misra (21) for the case $E_z = 0$. The elastic energy density (22) then becomes

$$u = \frac{a}{16g}\sum_l \left(\sum_{\alpha\beta} l_\alpha l_\beta e_{\alpha\beta}\right)^2 D^2\phi^l + \frac{3\hbar a}{128g}\left(\frac{5S}{m}\right)^{1/2}\Sigma_z \qquad \ldots\ldots(73)$$

where Σ_z is a series of sums due to the expansion of S and represents the effect of the zero-point energy. By comparing coefficients again with (23), Salter obtained the following expressions for the reduced elastic constants ($c_{ij}^* = \sigma^3 c_{ij}/\epsilon$)

$$\left.\begin{aligned} c_{11}^* &= \left(\frac{171{\cdot}2}{V^{*5}} - \frac{123{\cdot}0}{V^{*3}}\right) + \frac{1{\cdot}146\Lambda^*}{V^{*10/3}} f_{11}(V^{*2}) \\ c_{12}^* &= \left(\frac{84{\cdot}25}{V^{*5}} - \frac{54{\cdot}11}{V^{*3}}\right) + \frac{1{\cdot}146\Lambda^*}{V^{*10/3}} f_{12}(V^{*2}) \\ c_{44}^* &= \left(\frac{84{\cdot}25}{V^{*5}} - \frac{54{\cdot}11}{V^{*3}}\right) + \frac{9{\cdot}483\Lambda^*}{V^{*10/3}} f_{44}(V^{*2}) \end{aligned}\right\} \qquad \ldots\ldots(74)$$

where the terms containing $\Lambda^* f_{ij}$ are the quantum additions. The quantity ($c_{44}^* - c_{12}^*$), which is zero if the Cauchy relation is fulfilled, is seen to be zero only in the classical limit ($\Lambda^* \to 0$) and to increase for the lighter gases (Λ^* increasing). Using the Λ^* values corresponding to A, Kr, Ne, D_2 and a reduced equation of state, $p^* = f(V^*\Lambda^*)$ valid at 0°K, Salter obtained the variations of c_{12}^* and $(c_{44}^* - c_{12}^*)/c_{12}^*$ with p^* shown in figure 7. Typical values of $(c_{44} - c_{12})/c_{12}$ are only 7% for Kr at $p = 0$, but as much as 54% for Ne at $p = 450$ atm. Experimentally, these should be valid at temperatures below about $\Theta/5$, where the elastic constants become sensibly independent of temperature.

2.5.4. *Equations of state using model potentials.*

The first detailed calculation of an equation of state for the inert gas crystals was made by Herzfeld and Mayer (1934). They obtained the lattice energy U_0 as a function of r by summing an (exp, 6) interatomic potential (15) in which the constant ρ was made (*a*) 0·2091 Å, the value derived from quantum-mechanical calculations on the close approach of two neon atoms (Bleick and Mayer 1934) and (*b*) 0·345 Å, the value used by Born and Mayer (1932) in the repulsive energy term of the alkali halide crystals. The other constants (b, c) in the potential were obtained from a best fit to the experimental data on the second virial coefficient of argon gas. The summation for the attractive term ($-cr^{-6}$) was made over the whole crystal, using the Jones and Ingham (1925) values for S_6 (26), while that for the repulsive term was a limited series

$$\Sigma b \exp(-r/\rho) = 12b \exp(-r/\rho) \left\{ \begin{aligned} 1 + \tfrac{1}{2} \exp[-(\sqrt{2}-1)r/\rho] + 2 \exp[-(\sqrt{3}-1)r/\rho] \\ + \exp[-(\sqrt{4}-1)r/\rho] + \tfrac{4}{3} \exp[-(\sqrt{5}-1)r/\rho] \end{aligned} \right\}, \qquad \ldots\ldots(75)$$

thus allowing for a reasonable number of neighbours.

This lattice energy $U_0(V)$ was then substituted in the Debye equation of state (64), which may be written

$$p + U_0'(V) = -\frac{d}{dV}(\tfrac{9}{8}R\Theta_D) - E_D \frac{d \ln \Theta_D}{dV}, \qquad \ldots\ldots(76)$$

where

$$E_D = 9RT\left(\frac{T}{\Theta_D}\right)^3 \int_0^{\Theta_D/T} \left(\frac{x^3}{e^x - 1}\right) dx$$

is the Debye thermal energy. The parameter ($\nu_D = k\Theta_D/h$) was obtained in an involved calculation from the interatomic potential, and so from (76) Herzfeld and Mayer obtained two sets of isotherms for solid argon, corresponding to $\rho = 0{\cdot}2091$ Å and 0·345 Å.

Essentially the same procedure was followed by Kane (1939), who extended the calculations to Ne, Kr and Xe, but used new potential constants b, c obtained from the experimental values of lattice constant a_0 and sublimation energy L_0 of each crystal at 0°K. These constants refer to the minimum value of $E_0 = U_0 + E_z$, and not U_0 alone, as is made clear in figure 8 in the case of neon, where

$$a_0 = \sqrt{2}r_0 = 4{\cdot}52 \text{ Å}, \quad L_0 = 450 \text{ cal mole}^{-1}.$$

Kane's isotherms were very similar to those of Herzfeld and Mayer, his values for argon with $\rho = 0{\cdot}345$ Å being shown in figure 9. A characteristic feature of the isotherms of both authors is that beyond a certain temperature T_u the lattice becomes unstable mechanically. This is shown by a change in the slope of the isotherm (for example, $T_u = 76°\text{K}$ for $p = 0$ in figure 9), corresponding to a point of inflection in the curve of free energy as a function of volume. In such circumstances the lattice would disintegrate and this was thought to be associated with melting. A comparison of T_u with the known melting points T_m of these crystals (table 3) shows that in all cases they predict disintegration before melting naturally occurs. This instability is, in fact, a defect of the theory, which also predicts too rapid an increase with temperature of the expansivity and the specific heat C_p of these crystals.

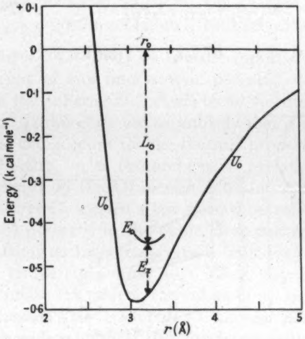


Figure 8. Lattice energy U_0, zero-point energy E_z and zero-point internal energy $E_0(=-L_0)$ for neon. (*After* Kane.)

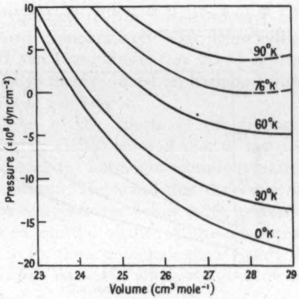


Figure 9. Isotherms for argon crystals based on the interatomic potential $\phi(r) = b\exp(-r/0{\cdot}345) - c/r^6$. (*After* Kane.)

Table 3. Instability and Melting Temperatures (Kane 1939)

	Ne	A	Kr	Xe	
$T_u(°\text{K})$	14	65	91	117	$(\rho = 0{\cdot}2091$ Å)
	21	76	104	144	$(\rho = 0{\cdot}345$ Å)
$T_m(°\text{K})$	24·5	83·8	116·0	161·3	

Kane's analysis is essentially a Debye theory, with $\Theta_D = \Theta_D(V)$ and as such involves the assumption (i) of harmonic vibrations, (ii) of a Debye frequency spectrum. Born (1943) pointed out that it was possible to dispense with the second of these assumptions fairly easily at high temperatures ($T \sim \Theta$), where classical statistics would apply. At high temperatures (β_j small), if we expand $\exp(-\beta_j)$, the free energy (55) becomes

$$F = U_0 + 3RT\ln\frac{\hbar}{kT} + kT\sum_{j=1}^{3n}\ln\omega_j, \qquad \ldots\ldots(77)$$

where the thermal energy is to be found directly from the frequency spectrum of the lattice. Now the sum of the squares of these frequencies is given by the second moment of the spectrum μ_2, which can be found from the secular determinant (43) by summations of a model potential over all neighbours, without solving the determinant for each frequency. In this way the frequency sum, from (77) and (43) is

$$\sum_{j=1}^{3n} \ln \omega_j = \tfrac{1}{2} \sum_j \ln \omega_j^2 = \tfrac{1}{2} \sum_j \ln |C_{\alpha\beta}(\mathbf{q})| = \tfrac{1}{2}\{\ln |C_{\alpha\beta}(\mathbf{q})|\}_{\text{Av}} \qquad \ldots\ldots(78)$$

where the average is to be taken over all wave vectors $\mathbf{q}$ in one reciprocal cell. The evaluation of $C_{\alpha\beta}$ involves sums of the general type

$$S_p^{p_1 p_2 p_3}(\mathbf{q}) = \sum_l \frac{l_1^{p_1} l_2^{p_2} l_3^{p_3}}{L^p} \exp[-2\pi i(l_1 q_1 + l_2 q_2 + l_3 q_3)] \qquad \ldots\ldots(79)$$

where $\mathbf{q}(q_1, q_2, q_3)$ is a point in reciprocal space, $\mathbf{l}(l_1 l_2 l_3)$ is a lattice vector and $p(p_1, p_2, p_3)$ are integers.

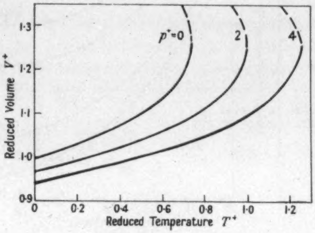


Figure 10. Isobars for cubic close-packed crystals, based on the interatomic potential $\phi(r) = A/r^{10} - B/r^6$. For argon, $p = 406p^+$ bars, $T = 110T^+$ °K, $V = 22{\cdot}6V^+$ cm³ mole⁻¹. (*After* Bradburn.)

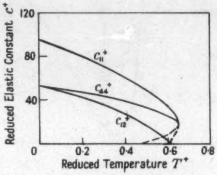


Figure 11. Elastic constants for cubic close-packed crystals at zero pressure, based on the interatomic potential $\phi(r) = A/r^{12} - B/r^6$. For argon, $c = 435c^+$ bars and $T = 118T^+$ °K. (*After* Gow.)

Such sums have been tabulated for the (12, 6) potential and the three cubic lattices by Born and Bradburn (1943). Bradburn (1943) used them to express the equation of state of the cubic close-packed lattice in terms of reduced units (p^+, V^+, T^+). She also showed that the sum (78) was equal to $\frac{1}{2}\ln|\{C_{\alpha\beta}(\mathbf{q}_i)\}_{\text{Av}}|$ (which is much easier to evaluate) within a few per cent. Hence she found reduced isotherms and isobars for the (12, 6), (10, 6) and (14, 6) potentials, the isobars for the (10, 6) potential being given in figure 10. Despite the improvement in the calculation of the lattice frequency, these show at temperatures of about Θ_D the same large expansivity and rapid instability which characterized Kane's isotherms. Gow (1944) completed this classical analysis of the cubic close-packed lattice by evaluating each reduced elastic constant as a function of p^+ and T^+ for a number of (n, m) potentials. Her results for the (12, 6) potential at $p = 0$ (figure 11) show that an initial linear decrease of each constant is followed by a rapid change as the instability temperature is approached.

Another equation of state for solid argon was that of Rice (1946), who assumed Grüneisen's law to be true with $\gamma = 4{\cdot}3$, the best fit to the experimental data of Simon and Kippert (1928) on $(\partial p/\partial T)_V$, Bridgman (1934) on the melting curve and Clusius (1936) on the specific heat. The Debye equation of state (64) at $p = 0$ then became

$$VU_0'(V) = -\frac{d\ln\Theta_D}{d\ln V}E_{\text{vib}} = 4{\cdot}3E_{\text{vib}} \qquad \ldots\ldots(80)$$

where Rice found $E_{\text{vib}} = 296{\cdot}7 + 29{\cdot}5(80-\Theta_D) - 0{\cdot}643(80-\Theta_D)^2$ cal mole^{-1} was in reasonable agreement with the Clusius specific heat. From these equations he obtained $U_0'(V)$ and so, by integration,

$$U_0 = -2007{\cdot}6 + 1410{\cdot}0x + 10710x^2 - 53947x^3 + 76500x^4 \text{ cal mole}^{-1} \quad \ldots\ldots(81)$$

where $x = (r - 3{\cdot}84)$ Å. This 'empirical' equation for the lattice energy, independent of any inter-atomic potential, shows comparatively large contributions from the anharmonic (x^3, x^4) terms for quite small amplitudes (say 0·1 Å) of vibration. It therefore seems unlikely that a Debye theory, based on harmonic vibrations, can account for the thermal properties of solid argon.

Finally, in a comprehensive investigation of the effects of the frequency spectrum on the Grüneisen constant γ, Barron (1955) showed that at low temperatures Grüneisen's law cannot be valid, even for harmonic vibrations under a (12, 6) potential in the cubic close-packed lattice. He found that the value of γ fell from its high temperature limit (γ_∞) at temperatures below $0{\cdot}3\Theta$, to reach a low temperature limit (γ_0) which depended on the shape of the interatomic potential. In table 4, the values of γ_∞ and γ_0 for three potentials show clearly that Grüneisen's law is favoured by a wide potential, especially a long-range attraction. Such wide potentials also give lower values of γ.

Table 4. Grüneisen's Parameter for Various Potentials in the Cubic Close-packed Lattice

Potential	(10, 4)	(8, 6)	(12, 6)
γ_0	2·19	2·16	2·82
γ_∞	2·22	2·27	2·96
$\gamma_\infty - \gamma_0$	0·03	0·11	0·14

2.6. *Anharmonicity in Ideal Crystals*

2.6.1. *Equation of state.*

The thermal energy of harmonic vibrations (§ 2.5) has been determined not only for the elementary models of Einstein, but also for the better model of Debye, and even in some cases (including the cubic close-packed lattice) for the correct frequency spectrum. The difficulties of extending such calculations to anharmonic vibrations are, however, formidable. The general problem has already been discussed in § 2.1, and may be expressed briefly as the solution of the wave equation (5) when terms higher than the harmonic (quadratic) one in the potential energy are not neglected. The anharmonic vibrations of the equilibrium lattice are therefore the solutions of

$$[T_n + U_2(u^2) + U_3(u^3) + U_4(u^4)]\,\chi(u) = \epsilon\chi(u) \qquad \ldots\ldots(82)$$

(compare equation (10)) and the thermal average of their eigenvalues ϵ give the anharmonic thermal energy and the anharmonic equation of state. So far only simple models have been used in deriving such anharmonic expressions.

The Einstein assumption for the harmonic lattice was that all the normal frequencies ν_j were the same, ν_E. This implies that each atom in the crystal vibrates independently of its neighbours, that is, as if they were at rest and just produced an isotropic, harmonic potential $(U-U_0)$ about each lattice point. Corner (1939) used such a model for calculating the zero-point energy ($E_z = \frac{3}{2}Nh\nu_E$) in the cubic close-packed lattice. Another application of this model was by Henkel (1955), who used it to solve equation (82) for the same lattice, where the cubic terms disappear by symmetry. He found the effective potential U for the oscillating atom by summing the interatomic potential, $\phi(r) = A/r^{10} - B/r^6$, over all neighbours for displacements with components x, y, z from a lattice point. This potential is of the form

$$U = U_0 + \alpha_1(x^2+y^2+z^2) + \alpha_2(x^4+y^4+z^4)$$

where α_1, α_2 are respectively the harmonic and anharmonic potential constants. Hence the wave equation (85) becomes

$$[T_n + \alpha_1(x^2+y^2+z^2) + \alpha_2(x^4+y^4+z^4)]\,\chi(u) = \epsilon\chi(u)$$

where the kinetic energy $T_n = -(\hbar^2/2m)\nabla^2$ and where we can allow

$$\chi(u) = \chi_1(x)\,\chi_2(y)\,\chi_3(z).$$

The wave equation is then separable into three equations of the form

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \alpha_1 x^2 + \alpha_2 x^4\right)\chi_1(x) = \epsilon\chi_1(x) \qquad \ldots\ldots(83)$$

and these can be solved by treating $\alpha_2 x^4$ as a perturbing potential for the otherwise harmonic oscillations.

In the absence of the perturbation, the eigenvalues (52) all have the same frequency $\nu_j = \nu_E$ and the free energy (55) becomes ($\beta_j = \beta$)

$$F = U_0 + 3NkT\beta/2 + 3NkT\ln(1-e^{-\beta}) \qquad \ldots\ldots(84)$$

where $\beta = h\nu_E/kT$ and $h\nu_E = \hbar(2\alpha_1/m)^{1/2}$. Henkel showed that for weak anharmonicity the corresponding free energy of the perturbed oscillators was

$$F = U_0 + 3NkT\frac{\delta}{2} + 3NkT\ln(1-e^{-\delta}) + 3N\Delta\frac{e^\delta+1}{(e^\delta-1)^2}$$
$$-\frac{3N\Delta^2}{2kT}\left\{\frac{e^{3\delta}+10e^{2\delta}+9e^\delta}{(e^\delta-1)^4}\right\} + \ldots \qquad \ldots\ldots(85)$$

where $\delta = h\nu/kT$, $h\nu = \hbar(2\alpha_1/m)^{1/2} + 3\hbar^2\alpha_2/4m\alpha_1$ and $\Delta = 3\hbar^2\alpha_2/4m\alpha_1$ is the anharmonic correction. For temperatures below 60°K this correction Δ is only about 5% of $h\nu$ and so the equation of state and specific heat derived from (85) are essentially Einstein functions with a number of correction terms. The specific heat

$$C_v = 3Nk\left[\frac{\delta^2 e^\delta}{(e^\delta-1)^2} - \frac{\Delta}{h\nu}K_1 + \left(\frac{\Delta}{h\nu}\right)^2 K_2\right] \qquad \ldots\ldots(86)$$

is seen to approach the Einstein form at low temperatures ($K_1, K_2 \to 0$) and so can only be expected to agree with experiment when the frequency spectrum has a negligible effect on the specific heat ($T > \Theta/5$, Katz 1951).

These anharmonic corrections were calculated by Henkel for argon in the range 10–60°K from a (10, 6) potential, whose constants were derived from the experimental values $U_0 = -2030$ cal mole^{-1} and $a_0 = 5{\cdot}40$ Å. Zucker (1956) repeated these calculations for argon using (12, 6) and (14, 6) potentials and extended them to neon and krypton. Since then, Domb and Zucker (1956) have recalculated the thermodynamic quantities $C_p, \alpha, \kappa_T, \gamma$ for temperatures in the range 10–80°K, using Henkel's method, with new, accurate, potential constants (see § 3.2). It will be seen below (§ 3.3) that these latest calculations are in better agreement with experiment than the earlier ones which we have discussed.

The only other detailed calculations of the effects of anharmonicity in crystals are those based on the one-dimensional, linear chain model. Damköhler (1935) first investigated anharmonicities in this model at classical temperatures ($T > \Theta$) and showed that any reasonable potential $\phi(y)$ between the particles in the chain (where y is the linear displacement) led to a specific heat at constant length (C_L) which decreased with temperature in the classical region. Dugdale and MacDonald (1954) extended this treatment to the next approximation for a Morse interaction potential

$$\phi(y) = D(e^{-2ay} - 2e^{-ay}) \qquad \ldots\ldots(87)$$

and obtained

$$C_L = Nk\left\{1 - \frac{kT}{8D} - \frac{9}{32}\left(\frac{kT}{D}\right)^2 - \ldots\right\} \qquad \ldots\ldots(88)$$

and

$$C_p = Nk\left\{1 + \frac{kT}{D} + \frac{15}{4}\left(\frac{kT}{D}\right)^2 + \ldots\right\} \qquad \ldots\ldots(89)$$

at zero pressure, confirming Damköhler's results. They also showed that the anharmonic corrections to the linear expansivity α_1 and compressibility κ_1 led to a difference between the Grüneisen constant (66), $\gamma_1 = \alpha_1 L/\kappa_1 C_L$, and that related to the characteristic frequency of the lattice ($\gamma_1 = -d\ln\nu_c/d\ln L$). In a later paper, MacDonald and Roy (1955) repeated this classical analysis using the (m, n) potential and obtained a linear expansivity,

$$\alpha_1 = \left(\frac{m+n+3}{mn\epsilon}\right)\frac{k}{2}. \qquad \ldots\ldots(90)$$

The assumptions made are those appropriate to the inert gas atoms. MacDonald and Roy suggested that the dimensionless quantity $\alpha\gamma T$ was a useful measure of the anharmonicity present in a crystal.

2.6.2. *Thermal conduction.*

In the harmonic approximation, the thermal energy of a non-metallic crystal is distributed in the $3n$ normal modes of its lattice vibrations. These quantized vibrations, or ***phonons***, may be treated as particles in a gas for the study of transport properties. The phonons in a harmonic crystal then correspond to the atoms in a rarefied gas, and there are no phonon–phonon collisions, but only phonon–wall collisions at the boundaries of the crystal. In these conditions, there is no thermal resistance inside the crystal and heat flows through it without a temperature gradient. There is also, as in a rarefied gas, no mechanism by which thermal equilibrium can be established throughout the crystal. Such conditions are not realized at classical temperatures, since the vibrations are anharmonic, and it is the presence of anharmonicity that produces thermal equilibrium and a finite thermal conductivity.

In this classical region, the phonon–phonon collisions have a mean free path λ in the phonon gas (Debye 1914) given by the kinetic relation

$$K = \tfrac{1}{3}\lambda v C \qquad \ldots\ldots(91)$$

where K is the thermal conductivity, v the average velocity of the phonons (that is, of sound, at long wavelengths) and C the specific heat per unit volume of the crystal. At classical temperatures v and C are roughly constant, so that K depends only on λ. This will decrease as $1/T$ as the temperature rises, since the number of phonons increases linearly with T in this region. No exact calculations of λ have been made, but Dugdale and MacDonald (1955) have proposed that

$$\lambda \simeq r_0/\alpha\gamma T, \qquad \ldots\ldots(92)$$

where r_0 is the nearest-neighbour distance in the crystal and $\alpha\gamma T$ is a dimensionless parameter that ' measures the anharmonicity ' of the lattice in the classical region.

At lower temperatures the thermal resistance varies not only with the density of the phonons, but depends also on the nature of the collision process (Peierls 1929). Normally, this is characterized by the collision of two phonons to form a third, where the conservation of energy gives $h\nu_1 + h\nu_2 = h\nu_3$ and the conservation of momentum $h\mathbf{q}_1 + h\mathbf{q}_2 + h\mathbf{q}_3 = 0$, where $\mathbf{q}$ is the wave-momentum vector in reciprocal space. The conservation of momentum is also satisfied if $\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3 = 2\pi/\mathbf{a}_n$, where $\mathbf{a}_n$ is a basic lattice vector (§ 2.3) and so $2\pi/a$ the reciprocal lattice constant. Such processes were called ' Umklapprozesse ' (or flip processes) by Peierls, who showed that they died out at very low temperatures ($T \ll \Theta$) approximately as $\exp(-\Theta/2T)$. Berman, Simon and Wilks (1951) first detected such Umklapp-resistance in diamond and synthetic sapphire crystals. These substances have very high Θ values, so that measurements of K at temperatures in the range $\Theta/10$ to $\Theta/20$ were possible.

At still lower temperatures, the thermal conductivity of an infinite crystal would go on increasing, but in all real crystals the mean free path of the phonons will be limited (as in molecular flow down a tube) by the boundaries of the crystal. Then the thermal conductivity (91) becomes $K = \frac{1}{3}LvC$, where L is now, for example, of the order of the diameter for a cylindrical specimen. This ' size effect ' was first detected by de Haas and Biermasz (1935) and explained qualitatively by Casimir (1938). Their experiments used good single crystals of the alkali halides, quartz, etc., but in polycrystalline material the grain boundaries also interfere with the phonons and so L becomes of the order of the grain size. Berman (1953) has reviewed a number of thermal conductivity measurements on dielectric crystals and Klemens (1956) has recently summarized the whole field of thermal conduction in solids.

The theory of thermal conduction in non-metallic crystals is simpler than that in metals, where both electron–phonon and phonon–phonon collisions occur. Even so, the three main effects due to anharmonicity, Umklapp processes and boundary scattering (figure 12) are only qualitatively understood. It should be possible to evaluate quantitatively the thermal conductivity of the ideal crystals, at least in reduced coordinates as a law of corresponding states. At present this is one of the unsolved problems of crystal dynamics.

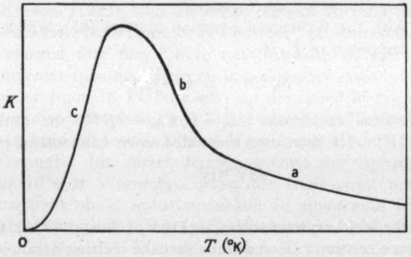


Figure 12. Thermal conductivity as a function of temperature in ideal crystals. Dominant effects : (*a*) anharmonicity, (*b*) Umklapp processes, (*c*) boundary scattering.

2.6.3. *Melting*.

There is at present no satisfactory kinetic theory of melting, but some progress has been made since Lennard-Jones and Devonshire (1939) first treated it as a type of order–disorder transition occurring in a crystal at a sufficiently high temperature. Their theory used the assumption of a central force and they did in fact find good agreement with the entropy and volume changes on melting for argon. Mott and Gurney (1939) reviewed melting theories in Volume V of these *Reports* and de Boer (1952) has since given an excellent review of developments in the theory up to 1952. Recently Rice (1954) has shown from thermodynamical principles that an infinite specific heat in an incompressible lattice becomes a first-order phase change in a compressible lattice and Domb (1956), applying this to the Ising and order–disorder models, concludes that ' even the crude model of melting of Lennard-Jones and Devonshire gives rise to a first-order transition at all pressures '. Certainly the melting process is no longer thought to be associated with the elastic instability of the harmonic lattice (Herzfeld and Mayer 1934) nor with the falling of its shear modulus to zero (Born 1939). Rather is it identified in the ideal crystals with the disappearance of long-range order : that is, a transition from regular to irregular ' close-packing ' of hard spheres.

There has recently been renewed interest in the study of the melting process at temperatures and pressures well above the critical point. The melting equation proposed by Simon in 1929

$$\frac{p_{\mathrm{m}}}{a} = \left(\frac{T_{\mathrm{m}}}{T_0}\right)^c - 1 \qquad \ldots\ldots(93)$$

where T_0 is the triple-point temperature (T_m at $p_m \simeq 0$), a is the internal pressure ($P_m \rightarrow a$ as $T_m \rightarrow 0$) and c is a constant for each substance, has been established experimentally for many elements. The most important measurements have been those on helium (Simon, Ruhemann and Edwards 1929, Holland, Huggill and Jones 1951, Mills and Grilly 1955) because very high values of p_m/a can be reached for this substance. There have been several theoretical derivations of Simon's relation, those by Domb (1951) and de Boer (1952) being based on the classical theory of Lennard-Jones and Devonshire (1939), while those of Salter (1954 b) and Fisher (1955) are more general.

Salter starts with the Grüneisen equation of state (65), which at the melting pressure p_m and temperature T_m is

$$p_m + a = \frac{\gamma}{V_m}(3RT_m) \qquad \ldots\ldots(94)$$

since under classical conditions ($E_z = 0, E_{vib} = 3RT$) the internal pressure $a = E_0'(V) = U_0'(V)$. He then uses the well-known Lindemann melting formula (Lindemann 1910)

$$\frac{MV^{2/3}\Theta^2}{T_m} = C, \qquad \ldots\ldots(95)$$

which holds for the ideal crystals with $C = (163^2)$ (Clusius 1936), to write (94) in the form of Simon's relation (93) and finds that the melting parameter

$$c = (6\gamma + 1)/(6\gamma - 2). \qquad \ldots\ldots(96)$$

This can be verified by measuring c for sufficiently high pressures of melting, if the constant γ is known from the high-temperature properties of the crystal. It can hardly be expected to hold for solid helium, however, for which the assumption $E_z = 0$ is not valid.

Nevertheless, solid helium does obey the Simon melting equation and the Lindemann melting formula, although not with the same constants as the other inert gas crystals. Domb (1952) has shown that the large zero-point energy, which prevents helium from solidifying at its 'normal melting temperature' (T_0 in equation (93) = 0·992°K), is far less effective for melting than the ordinary vibrational energy (E_{vib}). For example, at T_m = 13·1°K, the zero-point energy (170 cal mole^{-1}) is far larger than the thermal energy (6·4 cal mole^{-1}), but because the latter is concentrated at long wavelengths and the zero-point energy at short wavelengths, only the thermal energy breaks up the long-range order to cause melting. In a similar way, the mean amplitude of vibration (u_{RMS}) varies from 0·31a to 0·19a along the melting curve, whereas classically the Lindemann formula corresponds to the much smaller value $u_{RMS} \simeq 0{\cdot}1a$. Evidently helium melts as a 'quantum solid' and any theory of its melting curve must be sought in a strongly anharmonic treatment of its nuclear vibrations (Hooton 1955 b). For the ideal crystals the classical theory of Salter should be adequate.

§ 3. PROPERTIES OF SOLID ARGON

The verification of the theory of ideal crystals would require a knowledge of the properties of at least one such crystal and, preferably, of several similar substances. At present, the only ideal crystal for which accurate lattice measurements are available is solid argon. A good deal of information exists about solid helium but, as we have indicated, it is not suitable for our purpose (but see Domb and Dugdale 1957). The emphasis in the remainder of this report will therefore be on solid argon, existing data on the other inert gas crystals being considered where it is available. A summary of the known properties of the inert gas crystals, with estimates of their accuracy, will be found in the Appendix.

3.1. *Crystallization and Strength*

The first reported attempt to form a large piece of solid argon was that of de Smedt and Keesom (1925), who wished to measure its bulk density as a check on their lattice measurements from X-ray ' powder ' photographs. They did not succeed, and reported that they ' were not fortunate enough to solidify in a piezometer a sufficient quantity of argon in a compact mass '. We investigated the solidification of argon in 1951 by varying the speed of crystal growth from both the liquid and vapour phases and found that crystallization from the liquid produced a loose heap of crystals if it was rapid and a fairly transparent block if it was slow enough. But direct deposition from the vapour upon a surface cooled to about 70°K at a pressure below the triple-point pressure formed a completely transparent block, indistinguishable in appearance from glass. This technique was shown to produce a polycrystalline solid of grain size about 0·1 mm (Barker, Dobbs and Jones 1953) as estimated from X-ray diffraction photographs and from the attenuation of ultrasound. However, such argon grains grew rapidly at temperatures nearer the melting point (84°K), causing high attenuation of the ultrasound by grain-boundary (Rayleigh) scattering (Barker and Dobbs 1955, Guptill, Hoyt and Robinson 1955).

Stewart (1955) overcame the difficulties of making a compact specimen of solid argon in a very different way. His condensed solid was at first a loosely packed aggregate of crystals in a very small (~ 1 cm^3) steel cavity, but became a compact, elastic block when it was compressed to about 4000 atm with a steel piston. Although Stewart did not state the minimum pressure needed for compacting, it seems likely that considerably lower pressures might be sufficient, since he found that solid argon was very compressible and is easily extruded at high temperatures (table 5). Below 70°K it rapidly became less plastic and at 4°K both argon and krypton samples were shattered instead of being extruded, while neon could be extruded only with difficulty. At liquid helium temperatures, in fact, the only plastic solids are hydrogen, helium and their isotopes.

Table 5. Extrusion Pressures for the Rapid Flow of Cylindrical Solidified Gas Samples ($\frac{1}{2}$ in. × $\frac{1}{4}$ in. dia.) through $\frac{1}{8}$ in. hole (Stewart 1955)

Solid	A	A	A	Kr	Kr	Ne	H_2	D_2
Temperature (°K)	77	63	4	77	4	4	4	4
Pressure (10^6 dyn cm^{-2})	260	570	—	330	—	660	240	290

Recently, Stansfield (1956) has reported details of some attempts to make single crystals of argon by progressive freezing from the melt (Bridgman's method). He first cooled a whole bulb of liquid very slowly, but this resulted in the argon forming ' vapour snakes ' (Verschingel and Schiff 1954) and freezing loosely. However, the slow cooling from the pointed end of a bulb of liquid argon produced a transparent specimen with grain diameters of about 4 mm. These could be photographed and showed deep thermal etching (Chalmers, King and Shuttleworth 1948) at the grain boundaries and some surface markings on the grains. Stansfield also devised a technique by which a rod of argon could be released from the walls of its container and pulled *in situ.* In this way, he found the tensile strength at 75°K to be about 50 g wt mm^{-2}, that is, 5×10^6 dyn cm^{-2}. It is thus only a small fraction of the latent heat of vaporization per unit volume ($\sim 3 \times 10^9$ erg cm^{-3}) with which it would be compared in a simple theory (Polanyi 1921, Orowan 1934) of strength. By analogy with the properties of metals we thus suppose that the mechanical strength of solid argon is again severely limited by flaws in the material and does not represent the strength of the argon lattice. The tensile strength is also only a very small fraction of the shear modulus 9×10^9 dyn cm^{-2} (see table 7). As shown by Mackenzie (1949), this fraction should not be smaller than 1/30 for a perfect single crystal of argon. Stansfield's measurements were taken near the melting-point, where it is reasonable to assume that the weakness was due to plastic flow at the grain boundaries.

No measurements of any of the mechanical properties of solid argon at low temperatures have yet been reported, but Bowden and Rowe (1955) have investigated the hardness and frictional behaviour of polycrystalline krypton down to 25°K. At this temperature there was negligible creep and the indentation hardness was independent of the time of loading. Although the measured hardness increased at the lower temperatures, it reached a maximum of only 26×10^6 dyn cm^{-2} which, as the authors point out, is very low and comparable with that of butter on a cold day. Such hardness measurements refer to the heavily deformed material at the tips of the indenter and the values obtained are likely to be of the order of, but several times greater than, the shear stress for plastic flow in the undeformed material; this is roughly confirmed by comparison with Stansfield's results. The coefficient of friction μ with sliders of several materials was about 0·6 at 90°K and 0·75 at 30°K, showing that in the inert gas crystals the combination of small shear strength s with small hardness p produces friction ($\mu = s/p$) comparable with that of other solids (Bowden and Tabor 1950). Near the melting-point, this frictional resistance apparently disappeared, but this was due to local surface melting of the type familiar in skating and ski-ing.

3.2. *Zero-point Properties*

The theoretical investigations of the thermodynamic properties of the ideal crystals (§§ 2.5, 2.6) are based on model potentials (figure 4) of the general form (67) $\phi(r) = \epsilon f(r/\sigma)$, where ϵ, σ are the potential constants for a particular atom. These constants are best determined (Kane 1939, Corner 1939) from the interatomic distance r_0 and sublimation energy L_0 of the crystal at the absolute zero, by summing the interatomic potential $\phi(r)$ over all neighbours to find the lattice energy $U_0(r)$ and allowing for the zero-point energy $E_z(r)$ in a suitable manner (figure 8). A detailed comparison of the lattice theory with the properties of any one crystal is therefore only possible when its two constants (r_0, L_0) have been measured accurately.

The distance r_0 is best found from an extrapolation to 0°K of low-temperature X-ray measurements of the lattice constant ($a_0 = \sqrt{2} r_0$ for close-packed crystals).

Accurate lattice measurements have so far only been made for argon (§ 3.3.2) and lead to

$$\underline{a_0 = 5{\cdot}31 \pm 0{\cdot}01\ \text{Å}.}$$

The standard error in a_0 for the other inert gas crystals is much greater at present (see table I in Appendix), so that it is not yet possible to apply lattice theory to all these substances in order to obtain a precise test of the law of corresponding states (§ 2.5).

The sublimation energy L_0 is not measured directly but can be evaluated from the data of several experiments.† It is the total heat of the crystal minus that of its vapour, namely,

$$L_0 = \left(L_{\mathrm{T}} + \int_0^T C_{\mathrm{p}} dT\right) - (\tfrac{5}{2}RT + \Delta H) \qquad \ldots\ldots(97)$$

where, for convenience, T may be taken as the triple point. Then L_{T} is the heat of vaporization of the solid of specific heat C_{p}, and ΔH is the correction for change in total heat of the (real) gas as it expands from its vapour pressure to zero pressure (Joule–Kelvin effect). Clusius and Frank (1939) have measured the heat of vaporization of liquid argon at its boiling point (87·29°K) and corrected it to the triple point (83·78°K), obtaining $1572{\cdot}5 \pm 1{\cdot}5$ cal mole^{-1}. Adding this to the heat of fusion (Clusius 1936) of $280{\cdot}8 \pm 0{\cdot}4$ cal mole^{-1}, we find

$$L_{\mathrm{T}} = 1853{\cdot}3 \pm 1{\cdot}9 \text{ cal mole}^{-1}.$$

The mean specific heat curve (§ 3.3.1) gives $\int_0^T C_{\mathrm{p}} dT = 395{\cdot}7 \pm 10$ cal mole^{-1} and ΔH, calculated by an approximation due to Whalley and Schneider (1955), is 17·4 cal mole^{-1}. Thus the best experimental value of L_0 for argon is

$$\underline{L_0 = 1850 \pm 12 \text{ cal mole}^{-1}.}$$

Similar considerations determine the value of L_0 (see Appendix) for the other inert gas crystals, which, due largely to the work of Clusius *et al.* (1936–1939), are fairly well known.

These data (a_0, L_0) effectively fix the minimum of the interatomic potential, but its curvature (or shape) near the minimum can also be determined experimentally from the zero-point compressibility κ_0 and Grüneisen parameter γ_0. The former can be estimated from the existing high-temperature measurements (table 6) to be about $0{\cdot}41 \pm 0{\cdot}02$ ($\times 10^{-10}$ cm^2 dyn^{-1}). On classical theory (Fürth 1944), using an (n, m) potential, one obtains $\kappa_0 L_0 = 9V_0/mn$. The experimental values give $mn = 64$, so that a (10, 6) potential is a reasonable fit. Corner (1939) shows that, when zero-point energy is allowed for, a (12, 6) potential then fits best. This agrees both with the harmonic calculation of Salter (1954 a) using a (12, 6) potential, which gives, with the new potential constants $\epsilon = 169 \times 10^{-16}$ erg,

$$\sigma = 3{\cdot}402\ \text{Å},$$

$\kappa_0 = 0{\cdot}37$, and with the anharmonic calculation of Domb and Zucker (1956), using the same potential, from which $\kappa_0 = 0{\cdot}38$. A better test of the theory will be possible when measurements have been made of the compressibility of argon and krypton down to temperatures of 10°–20°K. Similarly, measurements of the expansivity at low temperatures are required before the values of γ_0 (table 4) can be compared with the experimental

---

† At 0°K, the heat of sublimation L_0 is identical with the internal energy $-E_0 = -U_0 + E_z$, at zero pressure.

$$\gamma = \alpha V / \kappa_S C_p. \qquad \ldots\ldots(66)$$

In this connection, Figgins, Jones and Riley (1956) have recently shown in experiments on aluminium that the expansivity should be measured at least down to temperatures of about $\Theta/15$ for an adequate test. For argon and krypton, these temperatures are in the liquid helium range, where no accurate measurements of lattice parameters have yet been reported.

3.3. *Thermodynamic Parameters*

The equations of state found theoretically for the ideal crystals (§ 2.5 and § 2.6) can be compared with experiment by studying the variation with temperature and pressure of the specific heat, expansivity and compressibility. In practice, the specific heat C_p and expansivity α are usually measured at 'zero' pressure as a function of temperature, but the compressibility can be studied as a function of both variables.

3.3.1. *Specific heat.*

It is well known that the specific heat (C_v, T) curves of the elements can be represented by the Debye model within a few percent, so that a comparison of experiment with lattice theory requires highly accurate measurements. For the ideal crystals, Clusius (1936) and Clusius and Riccoboni (1937) have measured C_p from 10°K to their melting points. Using the Nernst–Lindemann relation $C_p - C_v = 0{\cdot}0214 C_p^2 (T/T_m)$, they have calculated C_v at low temperatures (10°–15°K) and hence obtained the Θ_D values given in table 2. However, recent accurate measurements of C_p for argon from 3° to 20°K (R. W. Hill 1954, private communication) and from 16° to 35°K (Figgins 1955, private communication), whilst agreeing with one another within $\frac{1}{4}$%, differ from Clusius' values by as much as 5% at 20°K. Only these recent measurements below 35°K can, therefore, be used in a detailed comparison with lattice theory, but a mean curve (see Appendix) through the results of Hill and Figgins up to 35°K and of Clusius between 35° and 85°K is helpful in calculating some other quantities.

The frequency spectrum of the cubic close-packed lattice obtained by Leighton (1948) (§ 2.4) was used by him to form a Debye plot of Θ_D/β against T/β, where the temperature parameter $\beta = (h/\pi k)(\alpha/2m)^{1/2}$, α being the force constant between nearest neighbours. In the case of argon, the latter can be accurately evaluated from the new potential constants of Domb and Zucker (1956), who write

$$\phi(r) = A r^{-12} - B r^{-6} \qquad \ldots\ldots(98)$$

with $A = 1{\cdot}63 \times 10^{-7}$ Å^{12} erg, and $B = 1{\cdot}05 \times 10^{-10}$ Å^{6} erg. Then

$$\alpha = \phi''(r_0) = r^2 D^2 \phi + 2D\phi = 12 r_0^{-8}\{12 A r_0^{-6} - 3B\} = 1131$$

and so $\beta = 44{\cdot}6$°K. In this way, theoretical Debye curves for argon (figure 13) can be drawn, without using experimental values of the elastic constants, and compared with the Θ_D values of Hill and Figgins. It is clear from the figure that the frequency-sensitive values of $\Theta_D(T < 10°)$ show the rise in Θ_D predicted by Leighton for central-force interactions ($\gamma/\alpha = 0$) in the cubic close-packed lattice. Above this range the experimental points are evenly spread about $\Theta_\infty = 81°$, in excellent agreement with Leighton's result and the Domb and Salter expression (49), which corresponds to 80°, when all neighbours are considered (Domb and Zucker 1956).

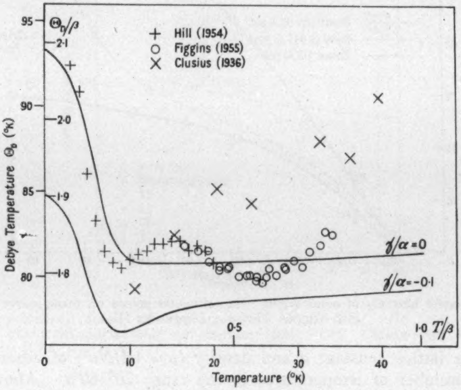


Figure 13. Debye plots for argon with central-force interactions ($\gamma/\alpha = 0$) and small anisotropy ($\gamma/\alpha = -0{\cdot}1$). (*After* Leighton.)

In deducing the experimental Θ_D values shown in figure 13, we have made use of the expansivity data of Dobbs *et al.* (1956) for the conversion $C_v = C_p/(1 + \alpha T\gamma)$, the values of $\alpha T\gamma$ being those given in the Appendix. Below 7°K, $\alpha T\gamma$ is zero within experimental error, but rises to 0·1 at 36°K. The variations of Θ_D in the temperature range 10°–35°K may, therefore, be slightly modified when more accurate values of $\alpha T\gamma$ are available. Above 35°K, the Clusius Θ_D values are too inaccurate to specify a temperature at which the deviations of Θ_D from Θ_∞ must be attributed to anharmonic vibrations. Clusius' specific heats are accurate enough, however, to show clearly the necessity of an anharmonic theory at high temperatures, as may be seen in figure 14 where the experimental C_p values (Appendix) are compared with the harmonic Debye theory of Kane (1939) and the anharmonic Einstein theory of Zucker (1957, private communication). Here the effect of the different frequency models is limited to low temperatures, while the potential model used by Kane (equation (15) with $\rho = 0{\cdot}345$ Å) is very similar to that (equation (17)) used by Zucker. It is worth noting, however, that C_v, though still rising at $T \sim \Theta$, does not rise above $3R$ (see Appendix III). There is thus no evidence here either of a fall in C_v due to anharmonicity, or of an 'additional' specific heat due to lattice defects.

3.3.2. *Density and expansivity.*

The density of solid argon has recently (Dobbs, Figgins, Jones, Piercey and Riley 1956) been measured to about 1 part in 1600 in the range 20° to 80°K by two independent methods. A low-temperature X-ray camera (Figgins *et al.* 1956) was adapted so that small, micro-crystalline specimens of argon could be formed at 20°K and photographed through thin windows in the radiation shields. In this way, the lattice constant *a* and density ($\rho = 4A/Na^3$) of argon was determined at a number of temperatures in the range 20°–60°K. Above this temperature, grain growth in the specimens (§ 3.1) made the X-ray method increasingly inaccurate, but the bulk density method, used by Barker and Dobbs (1955) as a rough check, was made precise and used in the range 60°–80°K. The two methods agreed within experimental error where they overlapped (figure 15), but differed appreciably from the earlier X-ray measurements of de Smedt and Keesom (1925) and Simon and von Simson (1924) at single temperatures. An extrapolation to the melting-point agreed with that deduced by Clusius and Weigand (1940) from observations of the melting curve and of the density of the liquid. Recently, Stewart (1956) has suggested, from some measurements of bulk density by Hinds, that the density of solid argon may be lower when it is formed from the vapour (A in figure 15) than when it is frozen from the liquid (B). It may be, however, that this discrepancy should be associated with the occurrence of fractional distillation, since his argon gas was only ‘ of the order of 99% pure ’, whilst that used by Dobbs *et al.* was of at least 99·999% purity.

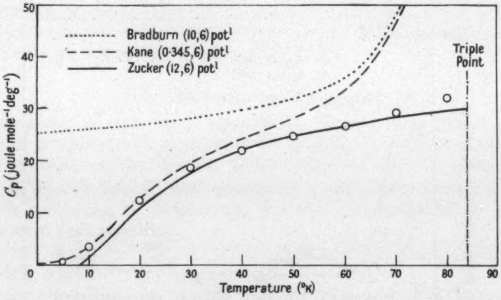


Figure 14. Specific heat C_p of solid argon. Experimental points on mean curve through Hill–Figgins–Clusius. (Appendix III.)

The density data (Appendix) are best compared with theory by considering the expansivities α derived from them by numerical differentiation (figure 16). The comparison shows clearly that an anharmonic theory is necessary above about 30°K for the expansivity, as well as for the specific heat (figure 14). At lower temperatures, the expansivity will be frequency-sensitive and the Einstein model used by Zucker is probably inadequate. Further expansivity measurements below 30°K are, however, necessary before this can be established with certainty. At classical temperatures ($T \sim 80°\text{K}$) the expansivity is about $16 \times 10^{-4}\ \text{deg}^{-1}$. This is much less than the value (36×10^{-4}) obtained by Macdonald and Roy in calculations based on an anharmonic linear chain (equation (90) with $n = 12$, $m = 6$ and $\epsilon = 169 \times 10^{-6}$ erg). It appears that the linear chain model is unsuitable for quantitative use. Their general result that, when the same potential parameters (n, m) can be used for a series of elements, the expansivities of these elements are inversely as their potential energies (ϵ) is true classically in the general case (Bradburn obtains $\alpha \propto 1/U_0$) of the harmonic crystal.

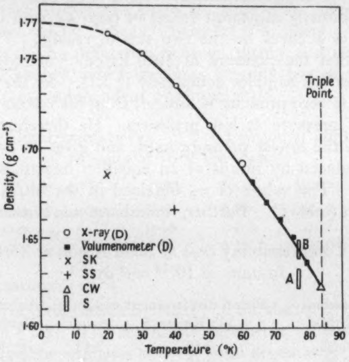


Figure 15. Experimental density of solid argon. D, Dobbs *et al.* 1956; SK, de Smedt and Keesom 1925; SS, Simon and von Simson 1924; CW, Clusius and Weigand 1940; S, Stewart 1956.

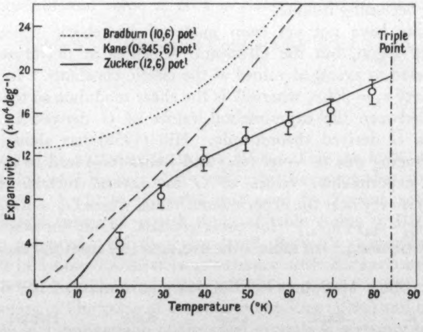


Figure 16. Expansivity of solid argon. Experimental points from Dobbs *et al.* 1956.

3.3.3. *Compressibility and elastic constants.*

The compressibility κ of solid argon has been measured both by the ultrasonic method as the adiabatic coefficient κ_S (Barker and Dobbs 1955) and by the piston-displacement method of Bridgman as the isothermal coefficient κ_T (Stewart 1955). The ratio of these coefficients, $\kappa_T/\kappa_S = C_p/C_v = (1+\alpha T\gamma)$, so that a comparison of the directly measured values of (say) κ_T with those deduced from $\kappa_S(1+\alpha T\gamma)$ serves as a check on the two sets of results. The good agreement obtained suggests that the pressure of 3000 Kg cm^{-2} employed by Stewart was sufficient to compact his samples completely (§ 3.1). As Stewart points out, an extrapolation of κ_T to zero pressure is difficult from his results, since this quantity varies rapidly with pressure at low pressures. He therefore quotes values for κ_T at 325 Kg cm^{-2}, the lowest pressure used, and gives an estimate for its value at zero pressure deduced by means of an equation due to Birch which fits the high-pressure data. The values of κ_T obtained in the ultrasonic measurements lie within these limits (table 6). Further, the anharmonic Einstein theory developed by Henkel (1955) with a (10, 6) potential and $a_0 = 5{\cdot}40$ Å is seen to be improved when based on a (12, 6) potential with $a_0 = 5{\cdot}31$ Å (Zucker 1956). The harmonic theories (§ 2.5) all predict instability at or near these high temperatures with the compressibility becoming infinite.

Table 6. Compressibility (κ_T) of Solid Argon at Zero Pressure in units of 10^{-10} cm^2 dyn^{-1}

Temp. (°K)	Ultrasonic Method	Piston displacement method p = 325 Kg cm^{-2}	$p = 0$	Anharmonic theory Henkel	Zucker
65	0·62	0·59	0·63	0·68	0·62
77	0·75	0·65	0·94	0·79	0·73

Measurements have not yet been made of the elastic constants of single crystals of solid argon, but the ultrasonic velocities in polycrystalline material are directly related to averaged values of the elastic constants. In particular, the transverse velocity $v_t = \sqrt{G/\rho}$, where G is the shear modulus, so that a comparison can be made between the experimental values of G derived in this way and averaged values $\bar{G}$ derived theoretically. Hill (1952) has shown that the two methods of averaging due to Voigt (G_V) and Reuss (G_R) lead to lower and upper limits for the experimental values of G for several metals. He finds that $\bar{G} = \frac{1}{2}(G_V + G_R)$ is very near the experimental value, where $G_V = \{(c_{11} - c_{12}) + 3c_{44}\}/5$ and $G_R = 5[4/(c_{11} - c_{12}) + 3/c_{44}]^{-1}$ for cubic crystals. Only for elastically isotropic crystals is $G_R = G_V = c_{44}$. In table 7 the averages ($\bar{G}$) from the classical harmonic theory of Gow, the harmonic Debye theory of Kane and the anharmonic Einstein theory of Henkel are compared with the values of G derived from the ultrasonic measurements. It will be seen that the experimental value falls within the range

Table 7. Shear Modulus G of Solid Argon in units of 10^{10} dyn cm^{-2}

°K	Expt.	Gow	Kane	Henkel
60	0·95	1·07	0·63	1·23
70	0·90	0·92	0·52	1·20

covered by the various theoretical averages and that for this modulus the harmonic theories give a reasonable answer.

A further comparison with theory would be provided if the variation of κ_T with pressure at low temperatures were known. Stewart (1956) has indeed measured κ_T up to 19 000 atm, but only at the high temperatures 65° and 77°K. Even so, above about 10 000 atm there is remarkable agreement between his values and those deduced by Salter (1954 a) for 0°K, using a (12, 6) potential, as is shown in table 8. The difference between theory and experiment is clearly due to the thermal energy, which becomes steadily less significant, and above 10 000 atm κ_T is almost independent of temperature.

Table 8. Compressibility (κ_T) of Solid Argon at High Pressures, in units of 10^{-10} cm^2 dyn^{-1}

Pressure (10^6 dyn cm^{-2})	1000	5000	10 000	15 000
Stewart κ_{65}	0·41	0·18	0·12	0·09
Stewart κ_{77}	0·48	0·185	0·12	0·09
Salter κ_0	0·30	0·16	0·105	0·08

3.3.4. *Grüneisen's constant.*

The importance of Grüneisen's constant γ in the theory of the crystal lattice has been emphasized in § 2.5. The recent measurements of density and expansivity have provided reasonable estimates of γ for solid argon as a function of temperature, and these values are shown in table 9, together with those derived by Domb and Zucker (1956), as later amended by them (Zucker 1957, private communication). It is apparent that at 40° and 60°K theory and experiment agree, but that at the extremes of temperature studied the experimental values are lower. The experimental value at 20°K is rather uncertain because of the large probable error in the expansivity below 30°K. That at 80°K depends on the specific heat value of Clusius, which may also be somewhat uncertain at this high temperature. A (12, 6) potential was used by Zucker and a comparison of the experimental values of γ with those of table 4 due to Barron (1955) also favours such a potential. On the other hand, Barron's suggestion that γ should fall by about 0·14 below $T \sim 0{\cdot}3\Theta$ is at variance with Zucker's calculations. We do not consider the experimental value at 20°K at present sufficiently accurate to decide these points. The value of 4·4 adopted by Rice (1946) and used by Murphy and Rice (1946) in a comparison of the ideal crystals is certainly too high.

Table 9. Grüneisen's Constant for Solid Argon

T (°K)	20	40	60	80
γ (exp.)	2	2·8	2·8	2·4
γ (theor.)	2·9	2·8	2·8	2·8

3.4. *Non-equilibrium Properties*

Crystals of pure argon are electrical insulators with a high breakdown voltage (~1 MV cm^{-1}), but they conduct heat owing to the anharmonicity of the thermal motions of their atoms (§ 2.6.2). An accurate verification of the basic ideas of the theory of thermal conduction in solids can only be made with good single crystals and, in so far as the theoretical conclusions apply to all dielectrics, is best undertaken with such substances as diamond, sapphire, quartz, etc., for which good mineral specimens are available. The value of measurements on argon is the hope that they can be explained quantitatively.

White and Woods (1956) recently obtained such information using a rod (3 in. × $\frac{1}{2}$ in. diameter) of polycrystalline argon formed in a thin-walled tube of 'Inconel' at 78°K. Their measurements of K extended from 2° to 25°K and are shown in figure 17. A comparison of this with figure 12 shows that at the lowest temperatures K must have been limited by the grain size of the specimen with $\lambda \sim 10^{-4}$ cm. This is a good deal smaller than the equilibrium grain size ($\sim 10^{-2}$ cm) at 78°K and suggests that the coherence of the solid must have been shattered by the severe thermal strains involved in cooling the solid from 78° to 7°K in a metallic tube. A large single crystal would undoubtedly have a much higher value of K at about 4°K. On the other hand, the 'high-temperature' variation of K with temperature corresponds roughly with the simple formula (92) of Dugdale and MacDonald (1955), as is shown in table 10.

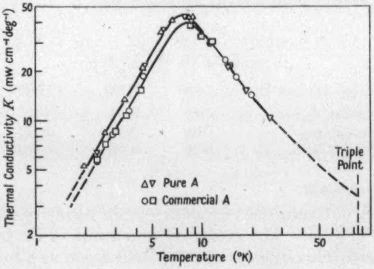


Figure 17. Experimental thermal conductivity of solid argon. (*After* White and Woods.)

Table 10. Thermal Conductivity of Solid Argon
(mw cm^{-1} deg^{-1})

T (°K)	10	15	20	25
K (exp.)	35	18	15	10
K (theor.)	100	65	50	40

The discrepancy in absolute magnitude of K is not large in view of the approximate nature of the theory, which contains no arbitrary constants. It is interesting to note that the values of K in figure 17, when extrapolated to the melting point, show solid argon to have about three times the conductivity of liquid argon (Uhlir 1952).

In its normal state, solid argon is an electrical insulator, and its dielectric properties depend only on its density, to a first approximation. The dielectric constant ϵ of solid argon can therefore be estimated from the data of Michels, ten Seldam and Overdijk (1951), who measured the molar polarizability p of argon gas up to a density of $1{\cdot}32\ \mathrm{g\,cm^{-3}}$. They showed that the Clausius–Mosotti relation

$$p = \frac{M}{\rho}\left(\frac{\epsilon-1}{\epsilon+2}\right) = \text{constant}$$

was valid within 1% up to the maximum density studied. Assuming that it remains valid up to $1{\cdot}77\ \mathrm{g\,cm^{-3}}$, the density of argon at 0°K, we obtain $\epsilon = 1{\cdot}67$. Hence the refractive index $n = \sqrt{\epsilon}$ of argon crystals for optical wavelengths is *c.* 1·29 at 0°K and falls to *c.* 1·26 at 84°K, the melting point. So far, no measurements of dielectric constant or refractive index have been reported for the ideal crystals. Nor has there been any study of their birefringence under stress, which would be expected to occur at temperatures well removed from the melting point.

However, some interesting observations on the use of solid argon as a crystal counter for nuclear radiations have recently been made by Hutchinson (1948). He irradiated a solid argon counter with γ-rays from RaC and applied an electric field of about $10\ \mathrm{kv\,cm^{-1}}$ to collect the electrons freed. Each photon gave rise to a larger pulse of current than would be expected from Compton collisions alone and Hutchinson suggests that there is considerable electron multiplication in solid argon. The effect is transient, however, and disappears with the accumulation of a positive space charge in the solid. In another investigation, Davidson and Larsh (1948) irradiated liquid argon with α-rays. In this case there could be no space charge, and they found that 0·7% O_2 in A quenched the pulses. Frerichs (1950) has suggested that all crystals which yield pulses under nuclear radiations belong to the class of crystal phosphors. Investigation of the optical absorption and of photoconductivity in solid argon is expected to be difficult, however, as the wavelengths fall in the Schumann region (1220–1800 Å).

3.5. *Melting Curve and Sublimation Curve*

The melting curves of He, Ne, A were studied by Simon, Ruhemann and Edwards (1929) in their classic investigation on melting up to pressures of 5000 atmospheres. Later, Robinson (1954) extended the curve for argon up to about 8000 atmospheres. All these results fit the Simon melting formula (93). In the later work on argon, the constants a and c of this formula have the values 3500 atmospheres and 1·16 respectively. Salter (1954 b) estimated that the melting parameter c should be related to the Grüneisen constant by (96); $c = (6\gamma+1)/(6\gamma-2)$. Taking $\gamma = 2{\cdot}8$ we find $c = 1{\cdot}20$. Too much importance should not be attached to this agreement since equally good agreement could be obtained by taking γ equal to, say, 4·5. However, the agreement is far better than for the case of solid helium for which $c = 1{\cdot}55$ and $\gamma = 2{\cdot}4$ (Dugdale and Simon 1953). Using de Boer's reduced coordinates, Domb (1952) has discussed the departures of these melting curves from a law of corresponding states.

The vapour pressures of solid and liquid argon have recently been determined accurately by Clark, Din, Robb, Michels, Wassenaar and Zwietering (1951) from 70°K up to the critical point. It is reasonable in this case to apply the vapour pressure equation

$$\ln p = i_0 - \frac{L_0}{RT} + \frac{5}{2}\ln T - \frac{1}{RT}\int_0^T \frac{1}{T^2}\int_0^T C_p\,dT.dT + X$$

where i_0 is the chemical constant, L_0 the latent heat of sublimation given by (97) and X is a small correction term to allow for the departure from ideality of the vapour. Using the best data (see Appendices) for L_0 and C_p, we find that the calculated value of p is less than the experimental value by 5–10% over the range 70°–84°K. The dominant term in the above expression is that containing L_0. However, because of the logarithmic form of the expression, a calculated value in better agreement with experiment would require much higher accuracy in the measurement of the latent heats, and to a lesser extent, of C_p.

3.6. *Solid Solution with Other Inert Gas Atoms*

Mixed crystals of the inert gas substances offer the possibility of providing material for the comparison of the various theories of solution in lattices which have been proposed. In particular, the calculation of the potential energy of a mixed crystal is made much simpler if the molecules concerned are of nearly equal size and have central forces. These conditions are fairly well met by argon and krypton, whose atomic diameters differ by only about 7%. Moreover, values of the interatomic potential for pairs of unlike inert gas atoms have already been evaluated (Buckingham 1937).

For these reasons a number of investigations have recently been made (Heastie 1955, Freeman and Halsey 1956, Figgins 1957) on mixtures of argon and krypton, which form solid solutions in all proportions. In order to estimate the critical temperature T_c below which separation into phases would occur, it is necessary to evaluate the energy gained in mixing an AB pair:

$$w = w_{AB} - \tfrac{1}{2}(w_{AA} + w_{BB}).$$

(If it appeared that w was negative, then a super-lattice might be formed at low temperatures for mixtures of simple ratio such as 1 : 1.) Considering only the van der Waals force and neglecting differences of size, we have $w_{AB} \sim \sqrt{(w_{AA} w_{BB})}$ since the polarizability is the chief factor of difference between A and B. This implies that $w > 0$, a conclusion which also follows from the detailed calculations of Buckingham. The influence of the repulsive potential would not be expected to modify this conclusion. Making use of the known heats of sublimation we can estimate that $w \sim 40$ cal mole^{-1} and that separation into phases might be expected between 10° and 40°K. Heastie (1955) and Freeman and Halsey (1956) have found positive deviations from Raoult's law, which imply $w > 0$, for solid solutions of A/Kr, and the latter authors estimate $T_c \sim 62$°K by the application of a rough treatment of regular solutions to their data. In fact, Figgins has found that the specific heat curve of a 1 : 1 solid solution of A/Kr lies midway between the curves for solid argon and krypton, respectively. There can be no separation into phases (nor significant change in the degree of order), at least down to 10°K, although the solid–liquid phase-equilibrium diagram found by Heastie for A/Kr mixtures has the form which suggests that T_c is not far below the solidus curve. Essentially similar conclusions have been reached in experiments on the Kr/Xe system.

It appears probable that such mixed crystals cannot be in equilibrium at temperatures appreciably below the solidus temperature—as suggested also by the earlier observations of Singleton and Halsey (1954) on the very slow rates of solution of argon in layers of krypton. Further information may be gained by x-ray studies of the state of order of these solutions.

§ 4. CONCLUSION

Since 1934, when Herzfeld and Mayer first made detailed calculations of the thermodynamic properties of solid argon, there have been many theoretical attempts to improve upon their results. These have been seriously handicapped by the lack of accurate experimental data, particularly on the lattice constants and elastic properties. This gap has been largely filled, with the result that it is now possible to choose between the various theories and to specify with some certainty the interatomic potential for a pair of argon atoms.

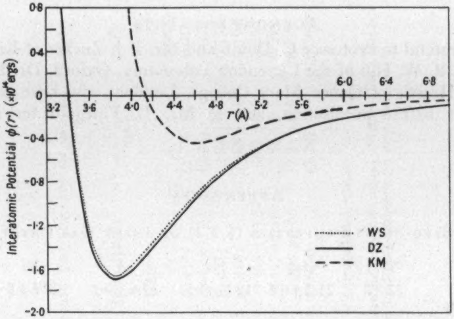


Figure 18. Interatomic potentials for a pair of argon atoms. WS, Whalley and Schneider, from gas data ; DZ, Domb and Zucker, from crystal data ; KM, Kunimune–Margenau, from quantum theory.

The potential which best fits the data on crystalline argon is a (12, 6) potential with constants evaluated by Domb and Zucker (equation (98)). This is almost identical with the potential which Whalley and Schneider (1955) obtained from their measurements on the second virial coefficient over a wide range of temperature. The position and depth of the energy minimum (see figure 18) are in fact equal for the two potentials within their statistical errors, being 3·819, 3·826 Å and 1·69, $1{\cdot}65 \times 10^{-14}$ erg respectively. The latter authors also fit an (exp, 6) potential to their data, but this gives noticeably poorer values for the crystal parameters a_0 and L_0. It does not appear therefore that the (exp, 6) form of the potential can be considered superior to the simpler (12, 6) form, in spite of its apparently sounder theoretical basis. The direct calculations of Kunimune (1950) on the repulsive interaction of argon atoms at short range gave results for three distances between 2 and 3·7 Å which he fitted to

$$515 \times 10^{-10} \exp(-r/0{\cdot}272)\,\mathrm{erg}.$$

It should be noted that his calculations referred to distances shorter than the distance between nearest neighbours in the argon lattice. The accuracy of direct calculation of the position and depth of the energy minimum is much less than that of the indirect estimate we have quoted.

The present position is likely to be further improved when experimental values of the expansivity, compressibility—and hence of Grüneisen's γ—are extended to lower temperatures, and perhaps to single crystals. There is also a great need for measurements of the optical properties of solid argon from which an accurate value of the polarizability might be obtained, and hence of the van der Waals potential constant. In the theoretical approach the next steps might be a better direct calculation of the interaction of argon atoms, and the development of an anharmonic theory using a better model than the Einstein approximation.

ACKNOWLEDGMENTS

We are grateful to Professor C. Domb and Mr. I. J. Zucker of King's College, London, Dr. R. W. Hill of the Clarendon Laboratory, Oxford, Dr. B. F. Figgins and Dr. R. Heastie of Queen Mary College, London, who have communicated results to us before publication, and to Mr. L. Finegold for assistance in calculations.

APPENDICES

I. ZERO-POINT PROPERTIES (§ 3.2) OF INERT GAS CRYSTALS

	^{3}He†	^{4}He‡	Ne	A	Kr	Xe
V_0 (cm^3 mole^{-1})	32 ± 2	$21{\cdot}2 \pm 0{\cdot}5$	$12{\cdot}5 \pm 1{\cdot}5$	$22{\cdot}6 \pm 0{\cdot}1$	$27{\cdot}6 \pm 0{\cdot}4$	35 ± 2
a_0 (Å)	—	—	$4{\cdot}35 \pm 0{\cdot}15$	$5{\cdot}31 \pm 0{\cdot}01$	$5{\cdot}68 \pm 0{\cdot}03$	$6{\cdot}1 \pm 0{\cdot}1$
L_0 (cal mole^{-1})	$4{\cdot}5 \pm 0{\cdot}5$	$12 \pm 0{\cdot}3$	450 ± 10	1850 ± 12	2590 ± 50	3830 ± 50
κ_0 (10^{-10} cm^2 dyn^{-1})	—	$23{\cdot}5 \pm 1{\cdot}0$	$1{\cdot}0 \pm 0{\cdot}1$	$0{\cdot}41 \pm 0{\cdot}02$	$0{\cdot}35 \pm 0{\cdot}05$	—

References: ^{3}He (Abraham *et al.* 1950, 1952, 1953, Kerr 1954); ^{4}He (Dugdale and Simon 1953); Ne (de Smedt *et al.* 1930, Clusius and Weigand 1940, Stewart 1956, Clusius 1936); A (see § 3.2); Kr (Dobbs 1956, Whalley and Schneider 1955, Stewart 1955); Xe (Natta and Nasini 1930, Clusius and Weigand 1940, Whalley and Schneider 1955).

† At 29 atm. ‡ At 25 atm.

II. CONSTANTS IN (12, 6) INTERATOMIC POTENTIALS (§ 2.5.2) FOR INERT GAS ATOMS

	^{3}He	^{4}He	Ne	A	Kr	Xe
ϵ($\times 10^{-16}$ erg)	14·11	14·11	49·2	169	230	311
σ (Å)	2·56	2·56	2·75	3·402	3·68	4·07
$\Lambda^* = h/\sigma\sqrt{m\epsilon}$	3·08	2·68	0·593	0·184	0·101	0·062

References: He, Ne (de Boer 1948); A (Domb and Zucker 1956); Kr, Xe (Whalley and Schneider 1955).

III. THERMODYNAMIC PROPERTIES OF SOLID ARGON

T (°K)	ρ (g cm^{-3})	V (cm^3 mole^{-1})	$10^4\alpha$ (deg^{-1} c)	C_p (joule mole^{-1} deg^{-1})	κ_S (cm^2 dyn^{-1}) $\times 10^{-10}$	$\gamma = \dfrac{\alpha V}{\kappa_S C_p}$	$\alpha T\gamma$	$C_v = \dfrac{C_p}{(1+\alpha T\gamma)}$ (joule mole^{-1} deg^{-1})	$\kappa_T = \kappa_s(1+\alpha T\gamma)$ (cm^2 dyn^{-1}) $\times 10^{-10}$
10	(1·769)	(22·58)	—	3·35 ± 0·03	—	—	(0·004)	3·34	—
20	1·764	22·64	4·0	12·55 ± 0·04	(0·414)	(1·77)	0·016	12·3	(0·420)
30	1·753	22·79	8·2	18·90 ± 0·04	(0·420)	(2·35)	0·058	17·8	(0·444)
40	1·736	23·01	11·7	21·9 ± 0·1	(0·431)	2·84 ± 0·5	0·133	19·3	(0·488)
50	1·714	23·30	13·5	24·6 ± 0·2	(0·444)	2·88 ± 0·5	0·194	20·6	(0·530)
60	1·689	23·65	14·9	26·8 ± 0·2	0·468 ± 0·05	2·81 ± 0·5	0·250	21·4	0·585
70	1·664	24·00	16·2	29·4 ± 0·2	0·512 ± 0·05	2·58 ± 0·5	0·292	22·8	0·662
80	1·636	24·42	18·0	32·1 ± 0·2	0·584 ± 0·05	2·35 ± 0·5	0·337	24·0	0·781
84	(1·623)	(24·61)	(18·9)	(33·0) ± 0·3	(0·620)	(2·28)	(0·362)	24·2	(0·844)
	± 0·001	± 0·015	± 1						

( ) = extrapolated value.

References: Dobbs *et al.* 1956 (ρ, V, α); R. W. Hill 1954 (private communication), Figgins 1955 (private communication), Clusius 1936 (C_p); Barker and Dobbs 1955 (κ_S, corrected for new density).

REFERENCES

ABRAHAM, B. M., OSBORNE, D. W., WEINSTOCK, B., 1950, *Phys. Rev.*, **80**, 366 ; 1952, *Ibid.*, **85**, 158 ; 1953, *Ibid.*, **89**, 787.

AXILROD, B. M., 1951, *J. Chem. Phys.*, **19**, 719.

AXILROD, B. M., and TELLER, E., 1943, *J. Chem. Phys.*, **11**, 299.

BARKER, J. R., and DOBBS, E. R., 1955, *Phil. Mag.*, **46**, 1069.

BARKER, J. R., DOBBS, E. R., and JONES, G. O., 1953, *Phil. Mag.*, **44**, 1182.

BARRON, T. H. K., 1955, *Phil. Mag.*, **46**, 720.

BARRON, T. H. K., and DOMB, C., 1955, *Proc. Roy. Soc.* A, **227**, 447.

BERMAN, R., 1953, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **2**, 103.

BERMAN, R., SIMON, F. E., and WILKS, J., 1951, *Nature, Lond.*, **168**, 277.

BLACKMAN, M., 1937, *Proc. Roy. Soc.* A, **159**, 416 ; 1941, *Rep. Progr. Phys.*, **8**, 11 (London : Physical Society).

BLEICK, W. F., and MAYER, J. E., 1934, *J. Chem. Phys.*, **2**, 252.

DE BOER, J., 1948, *Physica*, **14**, 139 ; 1952, *Proc. Roy. Soc.* A, **215**, 5.

DE BOER, J., and LUNBECK, R. J., 1948, *Physica*, **14**, 510.

DE BOER, J., and MICHELS, A., 1938, *Physica*, **5**, 945.

BORN, M., 1923, *Atomtheorie des festen Zustandes* (Leipzig : Teubner) ; 1939, *J. Chem. Phys.*, **7**, 591 ; 1940, *Proc. Camb. Phil. Soc.*, **36**, 160 ; 1942, *Ibid.*, **38**, 82 ; 1943, *Ibid.*, **39**, 100 ; 1944, *Ibid.*, **40**, 262 ; 1952, *Changements des phases*, 334 (Paris : Société de Chimie-Physique).

BORN, M., and BRADBURN, M., 1943, *Proc. Camb. Phil. Soc.*, **39**, 104.

BORN, M., and HUANG, K., 1954, *Dynamical Theory of Crystal Lattices* (Oxford : Clarendon Press).

BORN, M., and VON KÁRMÁN, T. H., 1913, *Z. Phys.*, **14**, 15.

BORN, M., and MAYER, J. E., 1932, *Z. Phys.*, **75**, 1.

BORN, M., and OPPENHEIMER, R., 1927, *Ann. Phys., Lpz.*, **84**, 457.

BOWDEN, F. P., and ROWE, G. W., 1955, *Proc. Roy. Soc.* A, **228**, 1.

BOWDEN, F. P., and TABOR, D., 1950, *The Friction and Lubrication of Solids* (Oxford : University Press).

BRADBURN, M., 1943, *Proc. Camb. Phil. Soc.*, **39**, 113.

BRIDGMAN, P. W., 1934, *Phys. Rev.*, **46**, 930.

BUCKINGHAM, R. A., 1937, *Proc. Roy. Soc.* A, **160**, 94.

CASIMIR, H. B. G., 1938, *Physica*, **5**, 495.

CHALMERS, B., KING, R., and SHUTTLEWORTH, R., 1948, *Proc. Roy. Soc.* A, **193**, 465.

CLARK, A. M., DIN, F., ROBB, J., MICHELS, A., WASSENAAR, T., and ZWIETERING, T., 1951, *Physica*, **17**, 876.

CLUSIUS, K., 1936, *Z. Phys. Chem.* B, **31**, 459.

CLUSIUS, K., and FRANK, A., 1939, *Z. Phys. Chem.* B, **42**, 395.

CLUSIUS, K., and RICCOBONI, L., 1937, *Z. Phys. Chem.* B, **38**, 81.

CLUSIUS, K., and WEIGAND, K., 1940, *Z. Phys. Chem.* B, **46**, 1.

CORNER, J., 1939, *Trans. Faraday Soc.*, **35**, 711.

DAMKÖHLER, G., 1935, *Ann. Phys., Lpz.*, **24**, 1.

DAVIDSON, N., and LARSH, A. E., 1948, *Phys. Rev.*, **74**, 220.

DEBYE, P., 1912, *Ann. Phys., Lpz.*, **39**, 789.

DEBYE, P., 1914, *Vortäge über die kinetische Theorie* (Leipzig : Teubner).

DOBBS, E. R., 1956, *Brit. J. Appl. Phys.*, **7**, 425.

DOBBS, E. R., FIGGINS, B. F., JONES, G. O., PIERCEY, D. C., RILEY, D. P., 1956, *Nature, Lond.*, **178**, 483.

DOMB, C., 1951, *Phil. Mag.*, **42**, 136 ; 1952, *Changements des phases*, **338** (Paris : Société de Chimie-Physique) ; 1956, *J. Chem. Phys.*, **25**, 783.

DOMB, C., and DUGDALE, J. S., 1957, *Progress in Low Temperature Physics*, **2**, 338 (Amsterdam : North-Holland).

DOMB, C., and SALTER, L., 1952, *Phil. Mag.*, **43**, 1083.

DOMB, C., and ZUCKER, I. J., 1956, *Nature, Lond.*, **178**, 484.

DUGDALE, J. S., and MACDONALD, D. K. C., 1954, *Phys. Rev.*, **96**, 57 ; 1955, *Ibid.*, **98**, 1751.

DUGDALE, J. S., and SIMON, F. E., 1953, *Proc. Roy. Soc.* A, **218**, 291.
EINSTEIN, A., 1907, *Ann. Phys., Lpz.*, **22**, 180, 800.
FIGGINS, B. F., 1957, *Ph.D. Thesis*, University of London.
FIGGINS, B. F., JONES, G. O., and RILEY, D. P., 1956, *Phil. Mag.*, **1**, 747.
FINE, P. C., 1939, *Phys. Rev.*, **55**, 355.
FISHER, I. Z., 1955, *Soviet Physics*, **1**, 280.
FREEMAN, M. P., and HALSEY, G. D., 1956, *Nature, Lond.*, **178**, 431.
FRERICHS, R., 1950, *J. Opt. Soc. Amer.*, **40**, 219.
FÜRTH, R., 1944, *Proc. Roy. Soc.* A, **183**, 87.
GOW, M. M., 1944, *Proc. Camb. Phil. Soc.*, **40**, 151.
GRÜNEISEN, E., 1926, *Handbuch der Physik*, **10**, 22 (Berlin : Springer).
GUGGENHEIM, E. A., 1945, *J. Chem. Phys.*, **13**, 255.
GUPTILL, E. W., HOYT, C. K., and ROBINSON, D. K., 1955, *Canad. J. Phys.*, **33**, 397.
DE HAAS, W. J., and BIERMASZ, T., 1935, *Physica*, **2**, 673.
HEASTIE, R., 1955, *Nature, Lond.*, **176**, 747.
HENKEL, J. H., 1955, *J. Chem. Phys.*, **23**, 681.
HERZFELD, K. F., and MAYER, M. G., 1934, *Phys. Rev.*, **46**, 995.
HILL, R., 1952, *Proc. Phys. Soc.* A, **65**, 349.
HOLLAND, F. A., HUGGILL, J. A. W., and JONES, G. O., 1951, *Proc. Roy. Soc.* A, **207**, 268.
HOOTON, D. J., 1955 a, *Phil. Mag.*, 46, 422 ; 1955 b, *Ibid.*, **46**, 485.
HUTCHINSON, G. W., 1948, *Nature, Lond.*, **162**, 610.
JANSEN, L., and DAWSON, J. M., 1955, *J. Chem. Phys.*, **23**, 482.
KANE, G., 1939, *J. Chem. Phys.*, **7**, 603.
KATZ, E., 1951, *J. Chem. Phys.*, **19**, 488.
KERR, E. C., 1954, *Phys. Rev.*, **96**, 551.
KIHARA, T., and KOBA, S., 1952, *J. Phys. Soc., Japan*, **7**, 348.
KLEMENS, P. G., 1956, *Handbuch der Physik*, **14**, 198 (Berlin : Springer).
KUNIMUNE, M., 1950, *J. Chem. Phys.*, **18**, 754.
LEDERMANN, W., 1944, *Proc. Roy. Soc.* A, **182**, 362.
LEIGHTON, R. B., 1948, *Rev. Mod. Phys.*, **20**, 165.
LENNARD-JONES, J. E., 1937, *Physica*, **4**, 941.
LENNARD-JONES, J. E., and DEVONSHIRE, A. F., 1939, *Proc. Roy. Soc.* A, **170**, 464.
LENNARD-JONES, J. E., and INGHAM, A. E., 1925, *Proc. Roy. Soc.* A, **107**, 636.
LINDEMANN, F. A., 1910, *Phys. Z.*, **11**, 609.
MACDONALD, D. K. C., and ROY, S. K., 1955, *Phys. Rev.*, **97**, 673.
MACKENZIE, J. K., 1949, *Ph.D. Thesis*, University of Bristol.
MARGENAU, H., 1939, *Rev. Mod. Phys.*, **11**, 1.
MICHELS, A., TEN SELDAM, C. A., and OVERDIJK, S. D. J., 1951, *Physica*, **17**, 781.
MIE, G., 1903, *Ann. Phys., Lpz.*, **11**, 657.
MILLS, R. L., and GRILLY, E. R., 1955, *Phys. Rev.*, **99**, 480.
MISRA, R. D., 1940, *Proc. Camb. Phil. Soc.*, **36**, 173.
MONTROLL, E. W., 1942, *J. Chem. Phys.*, **10**, 218 ; 1943, *Ibid.*, **11**, 481.
MONTROLL, E. W., and PEASLE, D. C., 1944, *J. Chem. Phys.*, **12**, 98.
MOTT, N. F., and GURNEY, R. W., 1939, *Rep. Progr. Phys.*, **5**, 46 (London : Physical Society).
MURPHY, G. W., and RICE, O. K., 1946, *J. Chem. Phys.*, **14**, 518.
NATTA, G., and NASINI, A. G., 1930, *Rendi. Atti. Accad. Lincei*, **11**, 1009.
OROWAN, E., 1934, *International Conference on Physics, London*, **2**, 81 (London : Physical Society).
PEIERLS, R., 1929, *Ann. Phys., Lpz.*, **3**, 1055 ; 1954, *Proc. Nat. Inst. Sci. India*, **20**, 121.
PITZER, K. S., 1939, *J. Chem. Phys.*, **7**, 583.
POLANYI, M., 1921, *Z. Phys.*, **7**, 323.
PRINS, J. A., DUMORE, J. M., and TJOAN, L. T., 1952, *Physica*, **18**, 307.
RICE, O. K., 1946, *J. Chem. Phys.*, **14**, 321 ; 1954, *Ibid.*, **22**, 1935.
ROBINSON, D. W., 1954, *Proc. Roy. Soc.* A, **225**, 393.
SALTER, L., 1954 a, *Phil. Mag.*, **45**, 360 ; 1954 b, *Ibid.*, **45**, 369.
SIMON, F. E., and KIPPERT, F., 1928, *Z. Phys. Chem.*, **135**, 113.

SIMON, F. E., RUHEMANN, M., and EDWARDS, W. A. M., 1929, *Z. Phys. Chem.* B, **6**, 62, 341.
SIMON, F. E., and VON SIMSON, C., 1924, *Z. Phys.*, **25**, 160.
SINGLETON, J. H., and HALSEY, G. D., 1954, *J. Chem. Phys.*, **58**, 1101.
SLATER, J. C., 1939, *Introduction to Chemical Physics* (New York : McGraw Hill).
DE SMEDT, J., and KEESOM, W. H., 1925, *Commun. Phys. Lab. Univ. Leiden*, **178b**.
DE SMEDT, J., KEESOM, W. H., and MOOY, H. H., 1930, *Commun. Phys. Lab. Univ. Leiden*, **203e**.
STANSFIELD, D., 1956, *Phil. Mag.*, **1**, 934.
STEWART, J. W., 1955, *Phys. Rev.*, **97**, 578 ; 1956, *J. Phys. Chem. Solids*, **1**, 146.
THIRRING, H., 1913, *Phys. Z.*, **14**, 867.
UHLIR, A., 1952, *J. Chem. Phys.*, **20**, 463.
VERSCHINGEL, R., and SCHIFF, H. I., 1954, *J. Chem. Phys.*, **22**, 723.
WHALLEY, E., and SCHNEIDER, W. G., 1955, *J. Chem. Phys.*, **23**, 1644.
WHITE, G. K., and WOODS, S. B., 1956, *Nature, Lond.*, **177**, 851.
ZUCKER, I. J., 1956, *J. Chem. Phys.*, **25**, 915.

CUMULATIVE AUTHOR INDEX

VOLUMES 16–20

Reports on Progress in Physics

PAGE
Ahrens, L. H. : Radioactive methods for determining geological age **19**, 80
Allen, C. W. : The physical condition of the solar corona **17**, 135
Ambler, E., and Hudson, R. P. : Magnetic cooling **18**, 251
Bacon, G. E., and Lonsdale, K. : Neutron diffraction **16**, 1
Bagguley, D. M. S., and Owen, J. : Microwave properties of solids **20**, 304
Blakemore, J. S., De Barr, A. E., and Gunn, J. B. : Semiconductor circuit elements **16**, 160
Bleaney, B., and Stevens, K. W. H. : Paramagnetic resonance **16**, 108
Blewett, J. P. : The proton synchrotron **19**, 37
Bouwkamp, C. J. : Diffraction theory **17**, 35
Bowers, K. D., and Owen, J. : Paramagnetic resonance—II **18**, 304
Briggs, B. H., and Spencer, M. : Horizontal movements in the ionosphere **17**, 245
Butler, C. C., with Rochester, G. D. : The new unstable cosmic-ray particles **16**, 364
Chalmers, J. Alan : Atmospheric electricity **17**, 101
Cole, G. H. A. : The kinetic theory of monatomic liquids at ordinary temperatures **19**, 1
Courtney-Pratt, J. S. : A review of the methods of high-speed photography **20**, 379
Craggs, J. D., and McDowell, C. A. : The ionization and dissociation of complex molecules by electron impact **18**, 374
Dalitz, R. H. : K-mesons and hyperons : their strong and weak interactions **20**, 163
Davies, R. O. : The macroscopic theory of irreversibility **19**, 326
De Barr, A. E., with Blakemore, J. S., and Gunn, J. B. : Semiconductor circuit elements **16**, 160
Dobbs, E. R., and Jones, G. O. : Theory and properties of solid argon **20**, 516
Fan, H. Y. : Infra-red absorption in semiconductors **19**, 107
Feldman, G. : A survey of field theory—Part V : High energy experiments **18**, 462
Gunn, J. B., with Blakemore, J. S., and De Barr, A. E. : Semiconductor circuit elements **16**, 160
Gunn, J. C. : Theory of radiation **18**, 127
ter Haar, D. : Theories of collective behaviour **20**, 130
Hudson, R. P., with Ambler, E. : Magnetic cooling **18**, 251
Jones, F. Llewellyn : Electrical discharges **16**, 216
Jones, G. O., with Dobbs, E. R. : Theory and properties of solid argon **20**, 516
Kay, H. F. : Electrostriction **18**, 230
Kinchin, G. H., and Pease, R. S. : The displacement of atoms in solids by radiation **18**, 1
Lee, E. W. : Magnetostriction and magnetomechanical effects **18**, 184
Lidiard, A. B. : Antiferromagnetism **17**, 201
Lonsdale, K., with Bacon, G. E. : Neutron diffraction **16**, 1
McCombie, C. W. : Fluctuation theory in physical measurements **16**, 266
McCrea, W. H. : Cosmology **16**, 321
McDowell, C. A., with Craggs, J. D. : The ionization and dissociation of complex molecules by electron impact **18**, 374
Matthews, P. T. : A survey of field theory—Part III : Interaction of quantized fields **18**, 441
Matthews, P. T. : A survey of field theory—Part IV : Properties of the π-meson **18**, 452
Mawardi, O. K. : Aero-thermoacoustics **19**, 156
Menzies, A. C. : Raman effect in solids **16**, 83
Mitchell, J. W. : Photographic sensitivity **20**, 433
Moffitt, W. : Atomic valence states and chemical binding **17**, 173

PAGE

Owen, J., with Bagguley, D. M. S. : Microwave properties of solids **20**, 304

Owen, J., with Bowers, K. D. : Paramagnetic resonance—II **18**, 304

Pease, R. S., with Kinchin, G. H. : The displacement of atoms in solids by radiation **18**, 1

Peierls, R. E. : A survey of field theory—Part I : General background **18**, 424

Peierls, R. E. : A survey of field theory—Part VI : Final remarks : the new particles **18**, 470

Price, B. T. : Ionization by relativistic particles **18**, 52

Pryce, M. H. L. : Nuclear shell structure **17**, 1

Raimes, S. : The theory of plasma oscillations in metals **20**, 1

Ratcliffe, J. A. : Some aspects of diffraction theory and their application to the ionosphere **19**, 188

Rochester, G. D., and Butler, C. C. : The new unstable cosmic-ray particles **16**, 364

Salam, A. : A survey of field theory—Part II : Fields and particles **18**, 433

Spencer, M., with Briggs, B. H. : Horizontal movements in the ionosphere **17**, 245

Stevens, K. W. H., with Bleaney, B. : Paramagnetic resonance **16**, 108

Symonds, J. L. : Methods of measuring strong magnetic fields **18**, 83

Tait, J. H. : Some topics in neutron diffusion theory **19**, 268

Taylor, A. E. : High-energy protons **20**, 86

Teller, Edward : Theory of origin of cosmic rays **17**, 154

Wilks, J. : The theory of liquid ^{4}He **20**, 38

Wooster, W. A. : Physical properties and atomic arrangements in crystals **16**, 62

Wynne, C. G. : New lens systems **19**, 298

CUMULATIVE SUBJECT INDEX

VOLUMES 16–20

Reports on Progress in Physics

PAGE
Absorption, infra-red, in semiconductors **19**, 107
Aero-thermoacoustics **19**, 156
Antiferromagnetism **17**, 201
Argon, solid, theory and properties of **20**, 516
Atmospheric electricity **17**, 101
Atomic arrangements in crystals, and physical properties **16**, 62
Atomic valence states and chemical binding **17**, 173
Atoms, displacement in solids by radiation **18**, 1
Binding, chemical, and atomic valence states **17**, 173
Collective behaviour, theories of **20**, 130
Corona, solar, physical condition of **17**, 135
Cosmic rays, new unstable particles **16**, 364
Cosmic rays, theory of origin **17**, 154
Cosmology **16**, 321
Crystals, physical properties and atomic arrangements in **16**, 62
Diffraction, neutron **16**, 1
Diffraction theory **17**, 35
Diffraction theory, some aspects, and application to ionosphere **19**, 188
Diffusion theory, neutron **19**, 268
Discharges, electrical **16**, 216
Dissociation, and ionization of complex molecules by electron impact **18**, 374
Electron impact, and ionization and dissociation of complex molecules **18**, 374
Electrostriction **18**, 230
Field theory, survey **18**, 424
Fluctuation theory in physical measurements **16**, 266
Geological age, radioactive methods of determination **19**, 80
Heat processes, and sound generation **19**, 156
Hyperons and K-mesons, their strong and weak interactions **20**, 163
Infra-red absorption in semiconductors **19**, 107
Ionization and dissociation of complex molecules by electron impact **18**, 374
Ionization by relativistic particles **18**, 52
Ionosphere, application of some aspects of diffraction theory **19**, 188
Ionosphere, horizontal movements in **17**, 245
Irreversibility, macroscopic theory of **19**, 326
Kinetic theory of monatomic liquids at ordinary temperatures **19**, 1
K-mesons and hyperons, their strong and weak interactions **20**, 163
Lens systems, new **19**, 298
Liquid helium, theory **20**, 38
Liquids, monatomic, kinetic theory, at ordinary temperatures **19**, 1
Magnetic cooling **18**, 251
Magnetic fields, strong, methods of measuring **18**, 83
Magnetostriction and magnetomechanical effects **18**, 184
Metals, theory of plasma oscillations in **20**, 1
Microwave properties of solids **20**, 304
Molecules, complex, ionization and dissociation by electron impact **18**, 374
Neutron diffraction **16**, 1
Neutron diffusion theory **19**, 268
Nuclear shell structure **17**, 1
Paramagnetic resonance **16**, 108

PAGE
Paramagnetic resonance—II **18,** 304
Particles, cosmic ray, new unstable **16,** 364
Particles, relativistic, ionization by **18,** 52
Photographic sensitivity **20,** 433
Photography, high-speed, review of methods **20,** 379
Physical measurements, fluctuation theory in **16,** 266
Plasma oscillations in metals, theory **20,** 1
Proton synchrotron **19,** 37
Protons, high-energy **20,** 86
Radiation, displacement of atoms in solids by **18,** 1
Radiation, theory of **18,** 127
Radioactive methods of determining geological age **19,** 80
Raman effect in solids **16,** 83
Resonance, paramagnetic, *see* Paramagnetic resonance
Semiconductor circuit elements **16,** 160
Semiconductors, infra-red absorption in **19,** 107
Solar corona, physical condition of **17,** 135
Solids, displacement of atoms in, by radiation **18,** 1
Solids, microwave properties of **20,** 304
Solids, Raman effect in **16,** 83
Sound, generation by turbulence and heat processes **19,** 156
Synchrotron, proton **19,** 37
Thermodynamics, macroscopic theory of irreversibility **19,** 326
Turbulence, and sound generation **19,** 156
Valence states, atomic and chemical binding **17,** 173

0
7

nd.
20
7